A contribution to online tool wear detection using deep learning methodology

by

Fatemeh AGHAZADEHKOUZEKONANI

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, 2020-05-25

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

M. Antoine Tahan, Thesis Supervisor
Mechanical Engineering Department, École de Technologie Supérieure

M. Marc Thomas, Co-supervisor
Mechanical Engineering Department, École de Technologie Supérieure

M. Gabriel J. Assaf, President of the Board of Examiners
Construction Engineering Department, École de Technologie Supérieure

M. Zhaoheng Liu, Member of the jury
Mechanical Engineering Department, École de Technologie Supérieure

M. Yann Landon, External Independent Examiner
Institut Clement Ader Université Fédérale Toulouse Midi-Pyrénées / Espace Clément Ader

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON "2020-04-23"

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my advisor, Prof. Antione Tahan, for giving me intellectual freedom in my research, moral support, his trust, scientific guidance throughout the course of this thesis, supporting my attendance at several conferences and helping me in generating and developing new ideas.

I would like to express my sincere gratitude to my co-advisor, Prof. Marc Thomas, for all his scientific contributions to make my PhD experience productive, supporting my thesis by his vast knowledge and skills, engaging me in new ideas and demanding a high quality of work in all my endeavors.

I am deeply grateful to all members of the jury for agreeing to read the manuscript and to participate in the defense of this thesis. I would like to thank all my friends and colleagues who helped me get through the difficult times. My special thanks to my friend Dr. Saman Bashbaghi and my colleagues at LIPPS and Dynamo laboratory including Mourad Kedadouche, Michel Drouin and all those people who have given me support and I did not have the chance to thank them. Lastly, and most importantly, I am grateful to my devoted and patient husband, Javad, my wonderful friend and companion who has inspired and supported me throughout this journey. I would also like to thank my family for their continued support and encouragement through my entire life. Without such a team behind me and their love I would not have finished this thesis.

Contribution à la surveillance en ligne de l'usure des outils de coupe à l'aide d'une méthodologie d'apprentissage en profondeur

Fatemeh AGHAZADEHKOUZEKONANI

RÉSUMÉ

La surveillance des conditions d'opérations dans les centres d'usinage est essentielle pour augmenter la productivité, améliorer la qualité et réduire les temps d'arrêt. L'usure des outils est une source courante des problèmes liées à l'usinage. Généralement, cette usure est due aux températures élevées et aux contraintes mécaniques imposées par le processus d'usinage. Par conséquent, l'industrie exige une surveillance fiable de la condition de l'outil pour répondre aux exigences de qualité et de productivité. La recherche proposée porte sur un système de surveillance en ligne de l'usure des outils dans les opérations de fraisage. Les signaux de force, de vibration et de courant sont utilisés comme indicateurs de panne pour développer un modèle robuste. Les principales étapes de la conception d'un système de surveillance intelligent sont l'acquisition du signal, son traitement, la modélisation de l'usure et la prise de décision. Toutes ces étapes sont abordées dans cette recherche. Une série d'expériences est menée avec la machine à commande numérique à haute vitesse Huron K2X10 dans les laboratoires LIPPS et Dynamo de l'ÉTS. Aussi, lors des travaux de développement et de validation, une base de données de référence pour l'usure des outils (NASA-Ames) est utilisée. Dans l'étape de traitement du signal, la transformation temps-fréquence est sélectionnée pour révéler simultanément les caractéristiques à la fois temporelles et fréquentielles des signaux. La transformation en ondelettes est utilisée pour transformer les signaux au domaine temps-fréquence. La méthode de soustraction spectrale est mise à profit par-dessus de la transformation en ondelettes afin de supprimer la partie stationnaire (stable) du signal et d'amplifier ainsi les signatures de défaut. Les récents développements des algorithmes d'apprentissage, notamment des méthodes d'apprentissage en profondeur, entraînent une amélioration significative du degré d'automatisation. Nous avons utilisé le réseau de neurones à convolution (CNN) comme algorithme d'apprentissage en profondeur pour la modélisation de l'usure des outils. La plupart des algorithmes courants d'apprentissage automatique sont également implémentés dans une approche comparative, et nous démontrons que le CNN surpasse les autres algorithmes. En outre, cette recherche porte sur la capacité d'évolution des algorithmes de surveillance afin de les rendre plus pratiques en introduisant un apprentissage par transfert en profondeur dans cette application. Malgré les avantages des algorithmes d'apprentissage automatique, l'un de leurs principaux inconvénients est qu'ils ont de grandes exigences en matière de données. Par exemple, un système de surveillance formé à partir des données d'une machine et d'un modèle spécifique n'est pas réutilisable sur une autre machine. Par conséquent, pour chaque machine et tâche, une quantité considérable de données de formation est requise. L'apprentissage par transfert fait référence à la réutilisation d'un modèle d'apprentissage sur une machine, qui est formé sur un problème pour un nouveau problème connexe (ou similaire). Cette méthode a été mise en œuvre et validée avec succès pour la surveillance de l'usure des outils, et il a été démontré qu'en exploitant le cadre de recherche proposé, il était possible d'obtenir de algorithmes de surveillance robustes, même avec de faibles quantités de données.

Mots-clés: Usure, Usinage, Surveillance, Apprentissage Par Transfert, Apprentissage en profondeur, Réseau de Neurones à Convolution.

A contribution to online tool wear detection using deep learning methodology

Fatemeh AGHAZADEHKOUZEKONANI

ABSTRACT

Condition monitoring is necessary in machining operation to increase productivity, improve quality and reduce downtime. Tool wear is one of the most common sources of machining problems which occurs due to high temperatures and forces of machining process. Therefore, industry demands reliable tool condition monitoring to address these requirements. This research investigates a robust on-line tool wear monitoring system in milling operation. Force, vibration and current signals are used as fault indicators to develop a robust model. The main steps of designing an intelligent monitoring system are signal acquisition, signal processing, wear modeling and decision making which all are tackled in different steps of this research.

A set of experiments are conducted with K2X10 Huron high speed CNC machine in LIPPS and Dynamo labs of ETS for validation of the research. Moreover, NASA-Ames tool wear benchmark data-set is used for further validation. In the signal processing step, time-frequency transformation is selected to reveal both time and frequency domain characteristics of the signals simultaneously. Wavelet packet transform as a well established algorithm is employed to transform the signals to time-frequency domain. Spectral subtraction method is leveraged on top of the wavelet transform for current signals to remove the steady state part of the signal and magnify fault signatures.

Recent developments in machine learning algorithms especially deep learning methods result in significant improvement in automation of various tasks in different industries. Therefore, we employed convolutional neural network (CNN) as powerful deep learning algorithm for modeling the tool wear. Most of the common machine learning algorithms are implemented as well in a comparative approach and it is shown that CNN outperforms the baseline algorithms.

Furthermore, this research focuses on scalability of the monitoring algorithms to make them more practical by introducing deep transfer learning in this application. Despite the advantages of machine learning algorithms, one of their main drawbacks is that they have large data requirements. For example, a monitoring system which is trained based on the data from an special machine and model is not reusable on another machine. Therefore, for each machine and task, considerable amount of training data is required. Transfer learning refers to reuse of a machine (deep) learning model which is trained on a problem for a related new problem. This method is successfully implemented and validated for tool wear monitoring and it is demonstrated that by leveraging the proposed framework of this research, robust monitoring algorithms can be achieved even with low amounts of data.

Keywords: Tool Wear, Condition Monitoring, Transfer Learning, Deep Learning, Convolutional neural network

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

AE	Acoustic Emission
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
ARMA	Autoregressive Moving Average
BPNN	Back Propagation Neural Networks
BSS	Blind Source Separation
CNN	Deep Convolutional Neural Networks
CWT	Continuous Wavelet Transform
DGAM	Distributed Gaussian ARTMAP
DWT	Discrete Wavelet Transform
ETS	École de Technologie Supérieure
FFT	Fast Fourier Transform
GPR	Gaussian Process Regression
GRU	Gated Recurrent Units
HMM	Hidden Markov Method
HRC	Hardness Rockwell C
KNN	Nearest Neighbors Regression
LDA	Linear Discriminant Analysis

LS-SVM	Least Squares Support Vector Machine
LSTM	Long-Short Term Memory
MLP	Multi-layer Perceptron
MSPCA	Multi-Scale Principal Component Analysis
NN	Neural Network
PCA	Principal Component Analysis
RBF	Radial Basis function
RMS	Root Mean Square
RNN	Recurrent Neural Networks
RUL	Remaining Useful Life
SAE	Sparse Autoencoder
SLD	Stability Lobe Diagram
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
SVR	Support Vector Regression
TCM	Tool Condition Monitoring
WT	Wavelet Transform

LISTE OF SYMBOLS AND UNITS OF MEASUREMENTS

a	Translation Factor
b	Additive Bias
D	Domain
h(t)	Short-Time Analysis Window
HRC	Rockwell C Hardness
k	Square Matrix
l	Index For Each Convolution Layer
M_j	Input Maps
mm/min	Milimeters Per Minute
$\mathscr{X}$	Feature Space
$P(X)$	Probability Distribution
rpm	Round Per Minute
u	Dilatation Factor
VB	Flank Wear
$\widehat{D}(jw)$	Steady State Component Of Spectrum
$\widehat{S}(jw)$	Fault Related Spectrum
X_i^{l-1}	Feature Map
Y(jw), S(jw), D(jw)	Frequency Domain Representation Of The Signal
y, s, d	Signals

β	Multiplicative Bias
μm	Micro Meter
$\psi(t)$	Base Wavelet

INTRODUCTION

Problem Statement

Milling is a machining operation that is among the most common activities in the manufacturing industry. The term machining refers to removal of unwanted material from the workpiece, so that a designed product with specific size, shape and surface quality can be manufactured. Manufacturing businesses currently have to cope with growing demand for increased productivity and production quality (Abellan-Nebot & Subirón (2010)). Tool wear is one of the most common faults of machining process and an important threat to those demands which can reduce efficiency and deteriorate the parts quality.

During the machining process, temperature and forces on the cutting tool edge are high. This may cause the cutting tool to gradually wears out and lose part of its material due to friction, thermal effects, abrasion, plastic deformation, diffusion, chemical wear, and grain-pullout, etc. There are several wear mechanism which may occur during the cutting process (Altintas (1992)) which can be listed as abrasion, adhesion, fatigue, crater, flank and chemical wear. Figure 0.1 depicts some of the tool wear types (TechniksUSA (2019)).

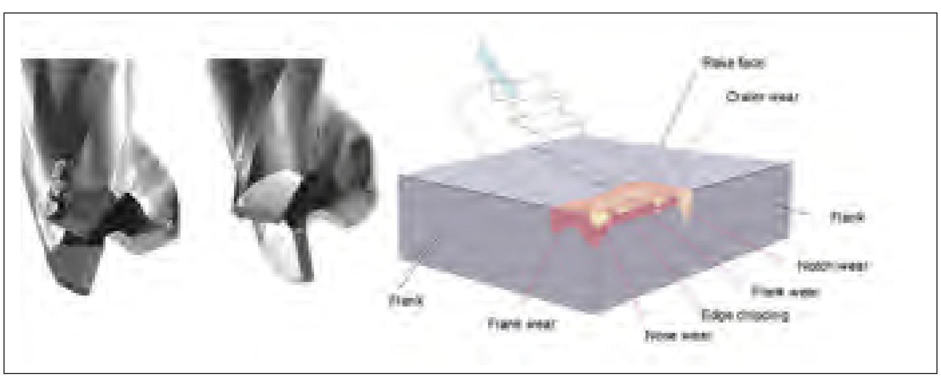

Figure 0.1 Examples of tool wear (TechniksUSA (2019))

This may cause damage to the tool, machine and workpiece and may lead to unscheduled downtime. Downtime refers to the time that a machine is not working which is a considerable factor in efficiency of a production line. The downtime due to tool faults and breakage is a major portion of total downtime in a machining environment. Moreover, to diminish the wear damage risks, cutting tools are usually replaced earlier than their maximum life which it will increase the production cost (Quintana & Ciurana (2011)). Instead, due to comparatively high prices of tools, industry is looking for approaches to maximize the tool usage while keeping low excessive wear and breakage risk.

Tool wear can also deteriorate surface finish (roughness) and is a limitation for production quality. Dimensional accuracy of the manufactured part, surface finish quality and tolerances are import requirements which can be endangered by tool wear.

Figure 0.2 summarizes some of the main problems of tool wears. This highlights the importance of automatically monitoring tool wear to ensure surface quality, maximum usage of tool without any damage to the material and changing the process variables such as feed rate and depth of cut for higher tool life.

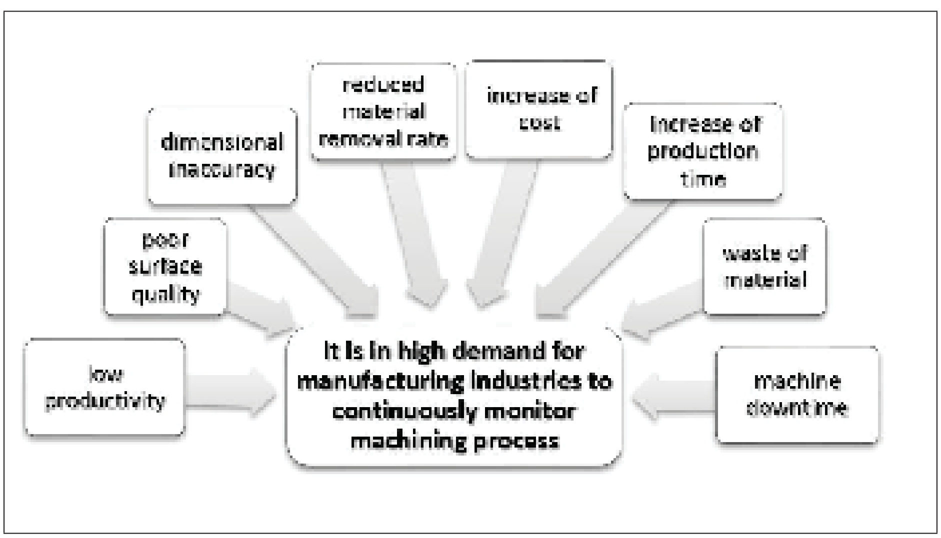

Figure 0.2   Importance of tool wear monitoring for the industry

There are many challenges in delivering such a system. The first requirement is that it should work on-line and real-time to guarantee immediate response when it is necessary. Moreover, it needs to detect tool wear early enough with high accuracy as well as minimum false alarms. Finally it needs to be scalable to be used with multiple machines and for different tasks accurately.

There are two main approaches in the literature for tool wear monitoring: direct and indirect methods. Direct methods such as laser, optical and ultrasonic monitoring are based on direct measurement of monitoring variable of interest. Although these methods are generally more reliable, they are not convenient for in-process use in a harsh manufacturing environment and they are still expensive or difficult to apply for online fault monitoring (Zhou & Xue (2018)).

Indirect methods estimate the machining faults by correlating it to auxiliary measured variables such as cutting force, vibration, acoustic emission, power signals and etc. (Zhou & Xue (2018)). Any deviation from normal situation can be an alert for possible faults in the systems. These methods are generally more economical and easier to apply in industrial environment. One of the challenges in using such approaches is that industrial environment is heavily noisy caused by other machines and environmental factors which will deteriorates quality of the signals. Moreover, the relationships between tool faults and those indicator signals are highly complex and depend also on other cutting parameters. Advanced signal processing algorithms as well as Artificial Intelligence (AI) methods can be leveraged to overcome some of these issues. Researchers have addressed the noisy nature of signals by various signal processing approaches in the literature. Time-frequency transformation is one of the advanced methodologies which has high potential in tackling the signal cleaning process. This approach considers both time domain and frequency domain characteristics of a signal simultaneously which empowers it to capture robust fault indicator from a faulty signal. This research benefited from spectral subtraction to address this issue.

Another challenge is the complex relations between the physical signals and tool wear (Benkedjouh *et al.* (2015)). Conventional machine learning methods can be used to solve this issue, however those usually require extensive signal manipulation before the machine learning step by the field experts. Another challenge is the scalability of such approaches as they require subnational data and training for each different task and machine. Those challenges are addressed in this research by proposing a deep transfer learning approach.

Research Objectives

The goal of this project is to develop a robust machining monitoring system for early detection of tool wear in machine tool cutters (e.g. cutting tools) to improve quality of the workpiece and productivity by reducing the loss of quality and downtime due to damage. This project will enable significant progress in the field of quality control because a machining monitoring system that meets these specific requirements is not available yet. Canadian aerospace manufacturers in particular are always looking for such intelligent systems. The focus of the research will be flank wear monitoring because of its importance in machining process(Dutta *et al.* (2016)).

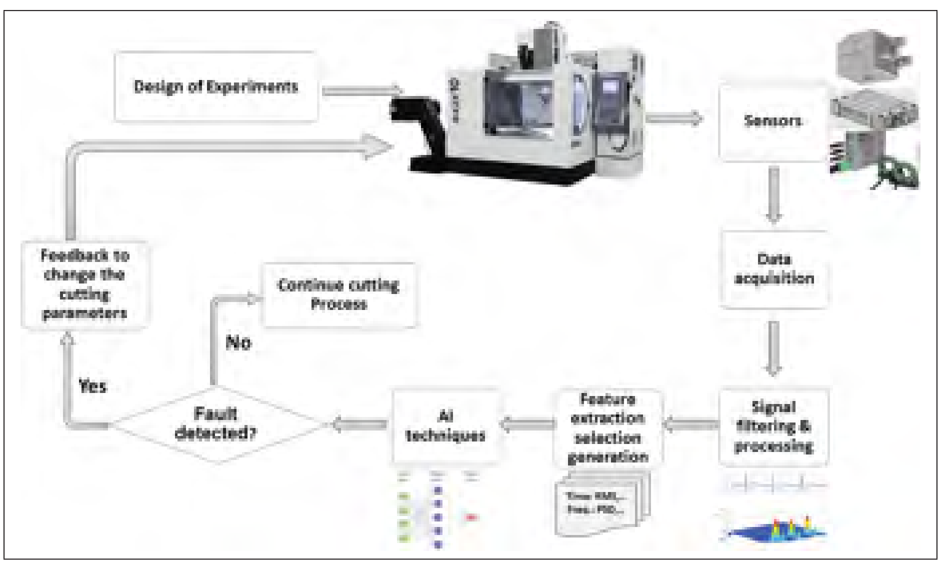

Figure 0.3   Proposed methodology scheme

Figure 0.3 represents this research's overall framework and methodology for developing such a system. The first step is to perform a set of experiments in order to obtain the required relevant data from the machining process using sensors and data acquisition technique. D2 high Speed Tool Steels with hardness of 60-62 HRC is considered as the workpiece material and a set of parameters are chosen for the experiments. Vibration, force and current sensors are used as practical and common sensors for signal acquisition. The focus of this study is to contribute to intelligent tool wear monitoring in three important steps. First processing and preparation of the signals is addressed using the advanced time-frequency techniques. Spectral subtraction is chosen to investigate the performance of this method in cleaning the signals. It is a method which was originally used for speech signal enhancement. In this method, a signal is considered a combination of noise and clean parts; It can be employed in fault diagnosis applications by removing the steady state and normal process spectrum from the new signals to obtain their anomalies and obtain fault signatures.

In the second part, the study focus is on leveraging the state of the art deep learning methodology to enhance modeling of the complex relationship between the signals and tool wear. It address the issue of handcrafted feature engineering requirements of previous methods and have the potential to improve accuracy and reliability of the monitoring systems. Deep convolutional neural networks (CNNs) is used for this task. The superior characteristic of this method is that the network learns data-driven filters to convert the data to features that describe the inputs and represent variables of interest inside the network which are usually performed separately in traditional methods (Lamraoui *et al.* (2015)). Therefore, higher accuracies can be achieved with less pre-processing by leveraging the data in big data scenarios.

Finally, the scalability issue of the machine learning techniques and high data requirements of new models are addressed by proposing transfer learning in this field as a novel methodology which enables us to leverage the information and knowledge gained from solving a problem in other systems.

Thesis Structure

This thesis is a thesis by article and its content is structured in 5 chapters, two published papers are presented in chapters 3 and 4 and one submitted paper is in 5. The literature review of related researches and state of the art in this field as well as research opportunities are covered in the Chapter 1. Chapter 2 describes the experimental protocol used in this project.

Chapter 3 investigates the signal processing and noise reduction step of tool condition monitoring and in particular, proposes spectral subtraction as an effective method for this task (Paper 1:Tool condition monitoring using spectral subtraction algorithm and artificial intelligence methods in milling process). Chapter 4 is devoted to exploring the use of deep learning methodologies (CNNS in particular) in tool condition monitoring (Paper 2: Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process). Chapter 5 provides an assessment of transfer learning to solve scalability and practicality issues of machine learning techniques in the field of tool condition monitoring (Paper 3: Tool condition monitoring method in milling process using deep transfer learning). Chapter 6 covers synthesis of the thesis and explains how a reliable condition monitoring system can be achieved by combining the results of different chapters. As represented in Figure 0.4, each chapter of this thesis is devoted to address one of the main aspects and problems of the intelligent tool condition monitoring by answering following questions:

Chapter 1: what are the practical sensors and data acquisition methods in the literature for tool condition monitoring?

Chapter 2: How to effectively design a set of experiments to acquire the required data for tool condition monitoring?

Chapter 3: what is an effective strategy to clean the signals coming from industrial environment and focus on the fault characteristics and signature rather than steady state part of the signal or environmental noises?

Chapter 4: How deep learning can contribute to increasing the accuracy and robustness of intelligent monitoring systems?

Chapter 5: How we can address some of the scalability issue and high data requirements of monitoring systems in similar tasks and machines for tool wear monitoring?

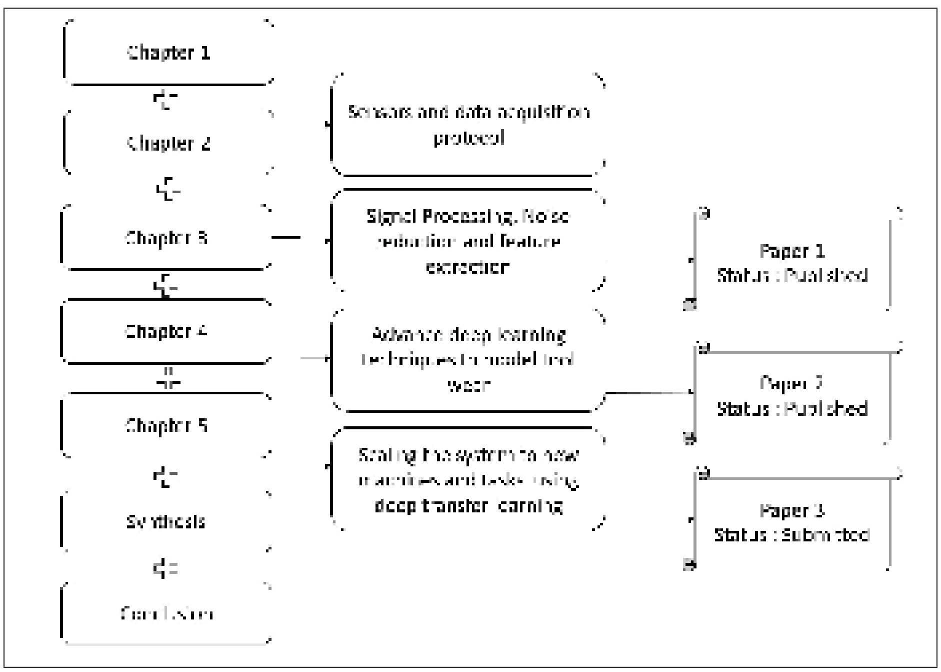

Figure 0.4 Requirements for intelligent tool condition monitoring

CHAPTER 1

LITERATURE REVIEW

There are two main approaches in the literature for tool wear monitoring: direct and indirect methods. Direct methods such as laser, optical and ultrasonic monitoring (Figure 1.1) are based on direct measurement of monitoring variable of interest. Although these methods are generally more reliable, they are not convenient for in-process use in a harsh manufacturing environment and they are still expensive or difficult to apply for online fault monitoring (Zhou & Xue (2018)).

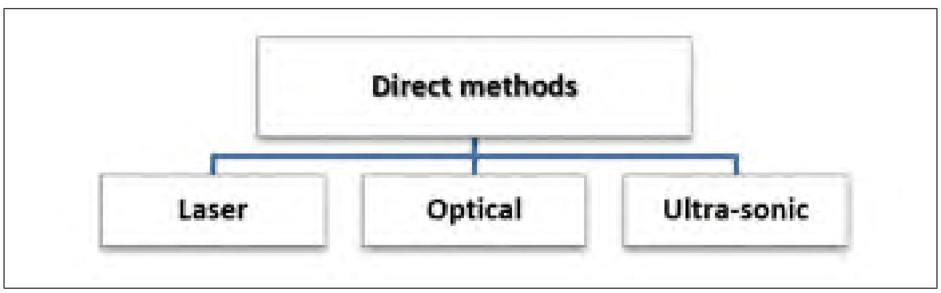

Figure 1.1 Direct methods for tool condition monitoring

Indirect methods estimate the machining faults by relating it to auxiliary measured variables such as, cutting force, temperature, vibration, acoustic emission, dimensions of the workpiece, etc. Any deviations from normal situation can be an alert for possible faults in the systems. These methods are generally more economical and easier to apply in industrial environment. Indirect methods use signal processing algorithms as well as artificial intelligence or mathematical modeling to detect the faults according to abnormality occurs to signals because of the faults (Figure 1.2).

1.1 Sensors and Data Acquisition

A sensor is a device which measures a physical quantity and converts its energy from one form to another. Selecting the most appropriate sensor is an important step in machining monitoring.

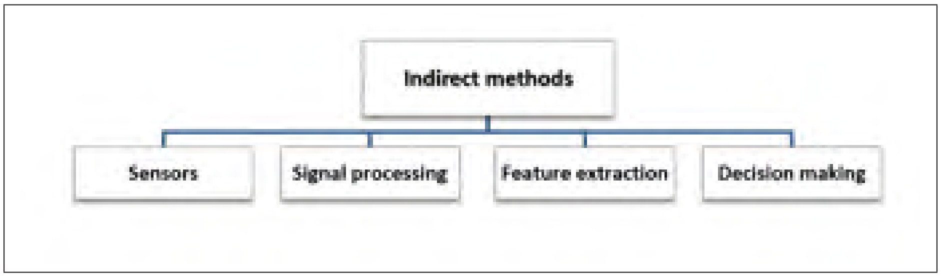

Figure 1.2 Indirect methods for tool condition monitoring

The ideal sensor for monitoring purpose is the one that is highly sensitive to the monitoring parameters and insensitive to the other process parameters. Reliability, sensor cost, ease of use in industrial environment also are among the other features must be taken into account (Abellan-Nebot & Subirón (2010)).

Various sensors are available for machining monitoring, which factors such as cost, accuracy and monitoring purpose can influence the sensor selection. Overall, four main types of sensors are employed in most of the monitoring machining and tool wear researches. Force and torque sensors, accelerometers, AE/sound sensors and motor power and current sensors. A multi-sensory system with sensor fusion of multiple sensors also is used in the literature for monitoring. Figure 1.3 represents the most commonly used sensors in this field based on their locations in the system (Nee (2015)).

Data acquisition refers to the process of sampling signals and converting the resulting samples into digital numeric values that can be manipulated by a computer. The signals should be filtered after signal acquisition within the range of the frequency response of the sensor. Signal sampling means the reduction of a continuous signal to a discrete signal. The sampling frequency should be high enough to enable a perfect signal reconstruction after sampling. There is a necessary relationship between the highest frequency contained in a signal and the minimum required sampling rate which is called the Nyquist sampling theorem. Based on this rule the sampling frequency should be at least twice the highest frequency contained in the signal being sampled (Siddhpura & Paurobally (2012), Hurmuzlu & Nwokah (2016)). This step may followed by some quality improvement methods and noise reductions.

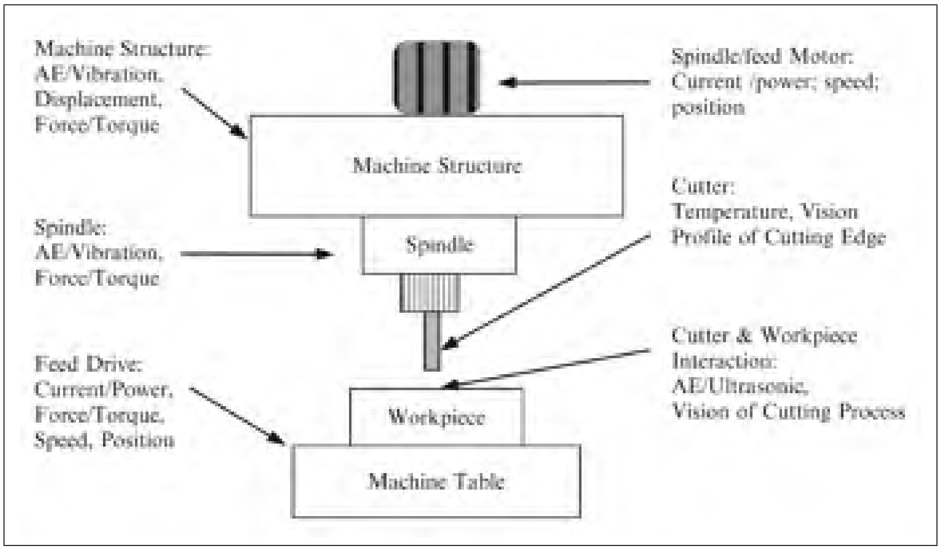

Figure 1.3 Most common sensors based on their locations (Nee (2015))

1.1.1 Cutting Force

Cutting force is considered as the signal which best describes the cutting process in terms of accuracy and fast responses to changes in cutting conditions. Cutting force signals have been widely used for monitoring purposes in all machining processes (Turning, Milling, Drilling, etc.). It is very sensible to tool faults and can reflect the fault very fast (Rehorn *et al.* (2005)). This signal can be used to monitor a wide range of machining variables such as tool wear, tool breakage/chipping, chatter vibration, and the quality and geometric profile of the cutting surface which makes it a perfect candidate to be used in a comprehensive machining monitoring system (Freyer *et al.* (2014), Zhang *et al.* (2012)). Torque signals are also used as another representation of force in machining monitoring application.

One of the drawbacks of using force signals is that they are highly dependent on cutting conditions, cutting material, workpiece material, etc. Therefore, an in increase in cutting force due to a fault is strongly dependent on other cutting variables. Furthermore, they are relatively expensive, difficult to install and their limited frequency response are among their limitations for industrial usage (Abellan-Nebot & Subirón (2010)), therefore they are not used in industry. As a research example, Wang *et al.* in a study employed force sensors and a distributed Gaussian ARTMAP (DGAM) network to solve non-uniformly distributed shapes and complex category boundaries problem for tool wear monitoring. They reported that, the classifier is insensitive to the noisy data and suitable for non-uniformly distributed data (Wang *et al.* (2013)).

1.1.2 Vibration

Interaction between workpiece and cutting tool produces vibration. Acceleration, velocity, or displacement sensors can measure vibration. Vibration is a significant factor affecting tool conditions, surface roughness, and dimensional accuracy in machining processes. As tool wears out or in case of tool breakage and other machining faults, amplitude and frequency of vibration signals changes which can be employed for monitoring purpose.

Many researches employed vibration signal for different applications of machining monitoring such as tool wear and breakage detection, surface quality control and chatter. The advantages of accelerometers are their simplicity and low cost. One of accelerometers practical problems is that it is sensitive to machining speed and it should remain within a specific range during the monitoring. Moreover, the placement of the sensor is a difficult task since it requires prior knowledge of the dynamic behavior of the machine tool(Delio *et al.* (1992)).

In an study by Gangadhar *et al.* condition of a single point cutting tool is monitored with help of the vibration signals acquired from an accelerometer (Gangadhar *et al.* (2014)). They obtained statistical features from vibration and the significant features were chosen using J48 algorithm. The accuracy of their monitoring system was 89.38% in distinguishing healthy and worn tools.

Lamraoui *et al.* (Lamraoui *et al.* (2015)) developed a chatter detection method in milling machine using vibration signals and cyclo-stationary analysis. They employed multi-band resonance filtering-envelope analysis. They claimed that chatter detection using dynamic cutting forces is not practical in industrial environments and used 3 accelerometers: 2 at spindle support in X and Y directions and 1 attached to work piece in cutting direction. Modal parameters are extracted using tap test and SLD is obtained in this research.

1.1.3 Acoustic Emission and Sound Sensors

Acoustic emission is defined as the energy release in the form of mechanical vibration from a material which undergoes stress. This stress may be generated by deformation, fracture, friction and thermal reactions of the tool, chips, workpiece, machine body, etc. The frequency range of AE signals is much higher than that of machine vibrations and environmental noises. A very high amplitude signal is generated due to tool breakage and fraction which makes this signal a very good candidate for tool breakage detection. However, there is still a debate in the literature about the usage of AE in detection of other faults such as wear. Many researchers used AE fault indicator signals in machining monitoring application (Prakash *et al.* (2014) Pai *et al.* (2012), Yen *et al.* (2013) Marinescu & Axinte (2008), Kosaraju *et al.* (2013)). AE sensors are easy to install and inexpensive, but they are sensitive to temperature and humidity. Since sound/AE signals are heavily dependent on process parameters, it is important to carry out the selection of signal processing methods and signal extraction techniques very carefully (Siddhpura & Paurobally (2012)).

Gowid *et al.* used AE signals and extracted FFT segment features of these signals for monitoring purposes. They suggested a high potential in using AE signals for fault detection (Gowid *et al.* (2015)).

1.1.4 Current, Power and Angular Encoder Sensors

Current, line voltage and phase shift measurement is an indirect way to sense the motor power which is proportional to cutting forces. Power sensors measure the spindle or axis drive power and current sensors measure the motor armature current. These sensors are not as accurate as dynamometers, however, they are economical and easy to install and they can be used as complementary information in monitoring systems. One of the major defects of these sensors is that they do not cover high frequency components of cutting forces and have slow response speed. Therefore, they are not appropriate for applications which need immediate response. Also, the sensitivity of power and current sensors are limited as they measure the total power required for the system not the portion that is used for cutting process itself (Bhuiyan & Choudhury (2014)). These sensors are employed for mainly tool condition monitoring applications (Rad *et al.* (2013), Ogedengbe *et al.* (2011), Ammouri & Hamade (2014))

Lamraoui *et al*. (Lamraoui *et al.* (2015)) proposed a monitoring algorithm based on Wiener-SVM approach by using motor current Signals. The total number of 14 experiments was used; ten of them corresponded to stable conditions and the other four with chatter phenomenon with aluminum. Total number of 23590 samples (one sample per one spindle revolution) is acquired which 17219 samples corresponded to stable case and the others 6371 samples included chatter phenomenon. In this paper the authors recommended that spindle current signal is a strong fault indicator and best time domain features to represent the chatter are Variance, RMS, Peak, and Clearance. They used three classification methods: SVM, Multi-layer perceptron (MLP) and Radial basis function (RBF) and suggested that the SVM approach provides better results versus MLP and RBF approaches

In another example, Altinas *et al*. (Altintas (1992)) employed current of the feed drive motors for milling process monitoring. The current and voltage limits in the amplifier as well as the friction in the feed drive assembly are included in the analysis. The paper discusses the performance of current signals as a cutting force measurement. The results show that tool failure in milling can be detected within one spindle revolution by adaptively filtering the average current signals at tooth passing periods. Imaouchen *et al*. investigated rolling element bearing defects, based on the wavelet packet decomposition (WPD) and the Hilbert transformusing motor current signal, which contains bearings fault-related frequency information Imaouchen *et al.* (2015). In another study, Proteau *et al*. leveraged specific cutting energy feature, defined as the amount of energy required to remove 1 cm^3 of material, to indirectly infer tool wear from power signals Proteau *et al.* (2019).

Signal from spindle integrated rotary encoder is also proven to be practical in chatter and tool wear detection (Lamraoui *et al.* (2014b)). Girardin *et al*. demonstrated that tool wear can be observed by monitoring variations in rotational frequency according to the different teeth of the tool using specific angular-sampling methodology (Girardin *et al.* (2010)). In another study, Lamraoui*et al*. leveraged encoder signal for re-sampling accelerometers signals in angular domain and used cyclo-stationary analysis for tool wear and chatter detection in slot milling operation of aluminum alloy (Lamraoui *et al.* (2014a)).

1.1.5 Sensor Fusion

Employing more than one sensor is useful to increase the reliability of the monitoring system. The sensitivity and the noise rejection of the sensed signal may alter by the change in cutting parameters, therefore, having multiple sensors with different characteristics helps to cope with this issue. Selection of sensors with complementary information is called sensor fusion. As an example, AE and force signals can be used effectively together as they are less correlated . However, using a dynamometer and a current sensor provide the same information with different level of accuracy, therefore, using them together is not sensor fusion (Bhuiyan & Choudhury (2014)). It should not be mixed with multi sensor monitoring which is another concept in which more than one sensor is used but not necessarily from different types.

There are many factors which affect sensor selection and fusion strategies. The cost-effectiveness and practicality of sensors, sensitiveness to variable of interest and being insensitive to other parameters in machining environment must be taken into consideration in selection of sensors. The sensors data should be complementary, and uniquely independent for most of the fusion applications. Another important parameter in designing a sensor fusion algorithm is the level of combining data. The data can be fused together at different levels such as sensor level, feature level and decision level (Bhuiyan & Choudhury (2014)).

There are several contributions in the field of machining monitoring using sensor fusion technique (Aliustaoglu *et al.* (2009), Wang *et al.* (2007), Cho *et al.* (2010)). For example, Cho *et al.*

combined information of force, vibration, acoustic emission, and spindle power sensor in time and frequency domain to develop an accurate and robust monitoring system. The workpiece material was 4340 steel and a multilayer-coated and multi-flute carbide end mill cutter is used as the cutting tool. Two different feature selection methods and three classifiers are evaluated and machine ensemble techniques are considered in the study (Cho *et al.* (2010)).

1.2 Signal Processing and Feature Extraction

Signals in industrial environment have high level of noise and may be not appropriate for directly using for monitoring purposes. Moreover, a signal has a wide range of information which some of them may be irrelevant to our monitoring parameters. Signal processing main function is to reduce the noise and irrelevant information of a signal and keep the principal components of a signal which have most correlation with the variable we are monitoring. As an example, cutting force signal may be filtered so that only relevant information to cutting such as the tooth-pass frequency is been kept. Another issue is transient mechanical events like breaking of a built-up edge, local variation in hardness over the work piece, etc. which may cause high frequency noises and signal oscillations which can be prevented by digital filtering. The features to represent the cutting process are extracted mainly from three domains:

- Time domain
- Frequency domain
- Time frequency domain

1.2.1 Time Domain Analysis

Time domain analysis is one of the most common methods in machining monitoring. Using this method has the advantage of simplicity. Generally, the original measured signal is in the time domain, in a format with the vertical axis as amplitude or voltage and the horizontal axis as time (Bhuiyan & Choudhury (2014)).

Analysis of the time domain signal in graphical format is very time consuming and not as informative because changes in the signal can happen for various reasons. Therefore, extracting representative features that can describe the signal adequately and maintain the relevant information about the process or tool conditions is necessary. Selecting the appropriate features highly depends on the signal nature itself and type of the cutting process and fault. This section investigates the appropriate features for each signal and presents some of the researches in tool wear monitoring using time domain analysis.

Two of the effective and simple descriptors for cutting-force signals are average and root mean square (RMS) values of cutting forces. Force ratios is another useful feature can also predict tool condition. Direction of the force is also important in prediction the faults. A large number of the researchers investigated three components of cutting forces (X, Y, and Z directions) to find the most significant descriptors for the monitoring variable. Mean and peak descriptors from cutting-force and power spectral amplitudes at the cutter-tooth frequency of cutting-force signals are also among the features suggested by the literature to have most correlation with the machining faults in time domain (Abellan-Nebot & Subirón (2010)).

For vibration signal, RMS or peak values of signals can also be used for fault detection as energy generated due to flank wear that increases the vibration magnitude (Yesilyurt & Ozturk (2007)). One of the drawbacks of using time domain analysis in vibration signals is that it is more sensitive to changing the cutting parameters than some faults.

Most of the time domain AE monitoring applications have used the RMS value of AE signals due to strong correlation of the AE RMS with the faults. After testing many different features of the AE signals, the average of the RMS signal, the average of the signal value, the burst rate and the pulse rate were most correlated with the faults.

Due to limited sensing bandwidth of current and power signals, they have been analyzed mainly in the time domain. Peak values, Kurtosis, mean and RMS of the current signal are among the features have been used for tool-wear estimation in milling. RMS value of spindle current in rough face milling operations correctly represented the cutting forces and the tool fracture was well distinguished from cutter run-out and transient cutting. Current features usually are used in a sensor fusion concept to improve the accuracy and reliability (Abellan-Nebot & Subirón (2010)).

1.2.2 Frequency Domain Analysis

A frequency domain analysis provides useful information from the signal in certain frequency range. In many machining monitoring systems the frequency spectrum of measured signals (such as vibration or cutting forces), carries a great deal of information that can be used to monitor the process. This method is widely used in the condition monitoring. Huang *et al*. (Huang *et al.* (2013)) performed vibration analysis in milling titanium alloy based on signal processing of cutting force using in Time frequency and Frequency domain. This paper mentions that Fourier transform has a drawback that non stationary transient information in time domain cannot be identified in frequency domain, and use of fast Fourier transform is much faster than discrete fourier transform and fourier transform.

1.2.3 Time-Frequency Domain Analysis

Although a lot of useful information can be extracted from time and frequency domain, it is still not sufficient for certain applications, especially in case of fault detection and due to the non-stationary nature of faulty signals. In the analysis of these signals, time-frequency analysis can identify the signal frequency components and at the same time reveal their time-varying features. Therefore, it is an effective tool to extract relevant information of a faulty signal for monitoring purposes (Feng *et al.* (2013)). Short time Fourier transform, Wavelet and empirical mode decomposition methods are among the widely used methods in this domain.

Short-Time Fourier Transform (STFT) represents the time-varying characteristics of a signal by adding a time variable to the traditional Fourier spectrum. This method assumes that in a short duration, the segmented signal is stationary due to minor changes. In implementation of this method, for higher frequency components, a shorter time window should be implemented and vice versa. This method is not suitable to analyze highly transient phenomena in signals, like impulses as the best time location and the frequency resolution cannot be obtained at the same time (Feng *et al.* (2013)). This method is best fitted to stationary signals. STFT formulation can be presented as follows:

$$STFT_x(t,v,h) = \int_{-\infty}^{+\infty} x(u)\, h^*(u-t)\, e^{-j2\pi vu} du \tag{1.1}$$

where $h(t)$ is a short-time analysis window localized around $t = 0$ and $v = 0$.

Wavelet transform (WT) is most widely used time frequency transformation for health condition monitoring systems. In this method, wavelets are used as the basis instead of sinusoidal functions. This method adds a scaling variable in addition to the time variable in the inner product transform which makes it an effective tool for transient signal analysis as well as time-frequency localization. It has a better time localization and a lower frequency resolution for higher frequency components. In contrast, the frequency resolution is higher for lower frequency components, while the time localization is worse. The continuous wavelet transform formulation is:

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\, \psi \frac{(u-t)}{a}\, du \tag{1.2}$$

where wavelet $\psi(u-t)/a$ is derived by dilating and translating the wavelet basis $\psi(t)$,and $1/\sqrt{a}$ is a normalization factor to maintain energy conservation and $a > 0$.

The Empirical Mode Decomposition (EMD) method is also another adaptive and efficient method for decomposing signals from high to low frequencies into intrinsic mode functions (IMFs) (Kedadouche *et al.* (2014)).

Rehorn *et al*. (Rehorn *et al.* (2006)) proposed a feature extraction method in time-frequency domain called selective regional correlation for machining faults monitoring. It is shown in their research that selective regional correlation can improve the resolution of monitoring. The performance of their approach is evaluated by short-time Fourier transform (STFT), continuous wavelet transform (CWT) and S-Transform and the results shows high accuracy and performance due to selective regional correlation.

1.2.4 Spectral Subtraction

Spectral subtraction method is one of the main algorithms proposed by Boll (Boll (1979)) for noise reduction in speech enhancement methods. This methods come on top of the normal time-frequency transformation signal to enhance the signal quality and reduce the noise, therefore it is a complimentary approach to time-frequency analysis and not its replacement. Assuming that noise is a stationary signal which almost does not change with time, the theory is based on the principle that noise spectrum can be estimated when speech is not present. And then it can be subtracted from the noisy speech signal which yields a clean speech signal spectrum. This method is originally used in speech enhancement to remove the effect of steady acoustic noise in the environment (Cho *et al.* (2010)).

This algorithm can be used in other domains including condition monitoring for noise reduction, and signifying the fault characteristics. El Bouchikhi *et al.* (Choqueuse *et al.* (2013)) proposed an algorithm for fault diagnosis of induction machine bearings using spectral subtraction method. In this study, stator current frequency response of the healthy machine is subtracted from the spectrum of machine current acquired signal to present better fault indicators. In another study by Martinez *et al.* spectral subtraction is used to detect rotor damages in induction motors based on the analysis of stray flux signals. It uses a spectral pattern recognition method by subtraction the spectrum of signal of the healthy motor from the power spectrum of the captured flux signals. The proposed algorithm is applied not only to detect adjacent bar breakages, but also nonadjacent broken bars. The results show the potential of this approach, which provides valuable information to detect rotor damages (Iglesias-Martinez *et al.* (2019)).

Wang *et al*. utilized spectral subtraction for fault diagnosis of a helical gearbox in combination with an Adaptive Empirical Wavelet Transform (AEWT). In this paper, the spectral subtraction technique is used to remove the partial noise of the vibration signal with strong noise disturbance, to enhance the fault information. Considering the benefits of this method in signifying the fault signature and removing the irrelevant steady state part of the signal, it is a great candidate to be investigated in tool condition monitoring for signal processing (Wang & Lee (2019)).

1.3 Decision Making Using Machine Learning

After the signal acquisition and processing step, features are extracted to represent system state. These features have complex and nonlinear relations with variable of interest in many cases. In order to predict machine state and make decision in machining monitoring such as tool wear detection, several researchers have employed artificial intelligence techniques. The applications of machine learning techniques in literature for intelligent tool condition monitoring is investigated in this section.

1.3.1 Conventional Machine Learning

Artificial intelligence techniques are able to make an estimation of the tool wear based on machining process conditions and features obtained from the sensory signals. Recently, many artificial intelligent techniques have been employed for classification of machine state in machine monitoring and tool fault detection, chatter detection and prediction such as artificial neural networks (ANNs), fuzzy logic systems, the hybridization of ANNs and fuzzy logic which is called neuro-fuzzy systems, BN Hidden Markov models, support vector machines and etc. (Siddhpura & Paurobally (2012), Stavropoulos *et al.* (2013))

Factors such as monitoring purposes, experimental data for modeling and previous knowledge of the process are important in choosing an AI technique. ANNs is a mathematical model which is inspired by the way biological nerve systems such as central nerve system of animals, in particular brain, process information. It can be used for applications like pattern recognition or data classification of signal features, through a learning process (Siddhpura & Paurobally (2012)). This method has numerous advantages such as the ability to learn from historical data, parallel computation which is suitable for real-time monitoring and ability to be used to extract and detect trends that are complex, to overcome the nonlinear difficulty, but requires large training data (Bishop *et al.* (1995), Rangwala & Dornfeld (1990), Lamraoui *et al.* (2015)). Neural networks architecture includes an input layer, some hidden layers and an output layer. Figure 1.4 depicts ANNs' generic architecture.

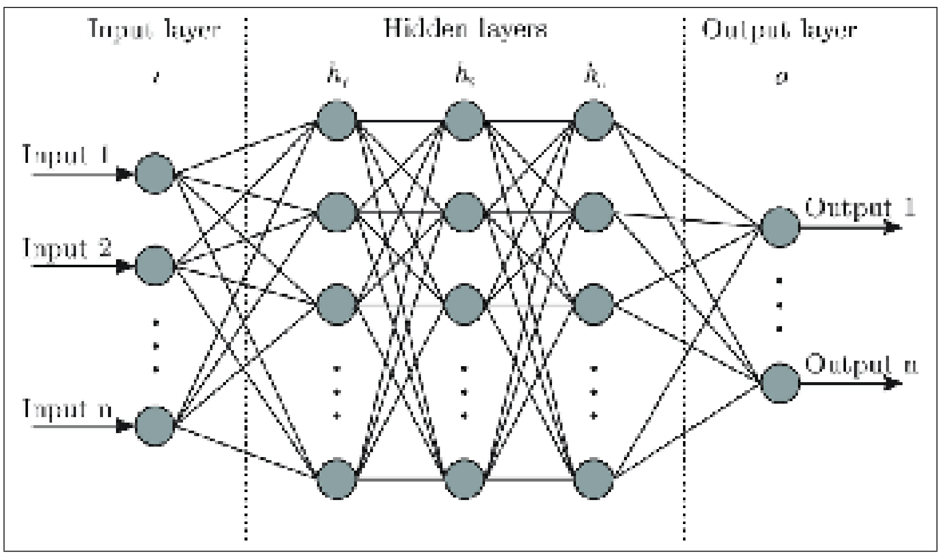

Figure 1.4 ANNs architecture (Bre *et al.* (2018))

Numerous studies leveraged neural networks in machining monitoring and wear detection. Tansel (Tansel (1992)) used neural network to simulate the cylindrical turning of long slender bars using vibration signals. The model used two neural networks to estimate the future vibration characteristics of system. The developed neural networks are capable of identifying 98% of the harmonic signals with over 90% certainty and with less than 5% error. Lamraoui *et al*. (Lamraoui *et al.* (2015)) used neural networks for chatter detection in milling machines using the acceleration signals obtained from 3 accelerometers and compared two neural network methods, radial basis function and multi-layer perceptron. They concluded that ANN is an adequate method for chatter detection as it has a non-stationary characteristic and both MLP and RBF have acceptable results regardless of depth of cut or cutting speed.

The application of Fuzzy systems is when the experimental data set consists of a low number of samples. In these systems, part of the model is developed using previous knowledge, so they will be used where there is enough knowledge from the process and this knowledge is intended to be added into the model. One of other applications is where the inverse problem should be solved. Generally, the accuracy of these systems is less than ANN (Abellan-Nebot & Subirón (2010)).

Neuro-fuzzy systems are a hybridization of ANN and fuzzy systems. Therefore, the application is like both application of ANN and fuzzy system. They are used when previous knowledge needed to be added to system and hidden knowledge from experimental data should be extracted (Solatian *et al.* (2012)).

Adaptive Neuro-Fuzzy Inference System (ANFIS) is a combination of fuzzy systems and neural networks to capture the advantages of both. This method generates and optimizes the fuzzy rule sets and parameters of membership function by training of the fuzzy interference systems using artificial neural networks (Solatian *et al.* (2012)).

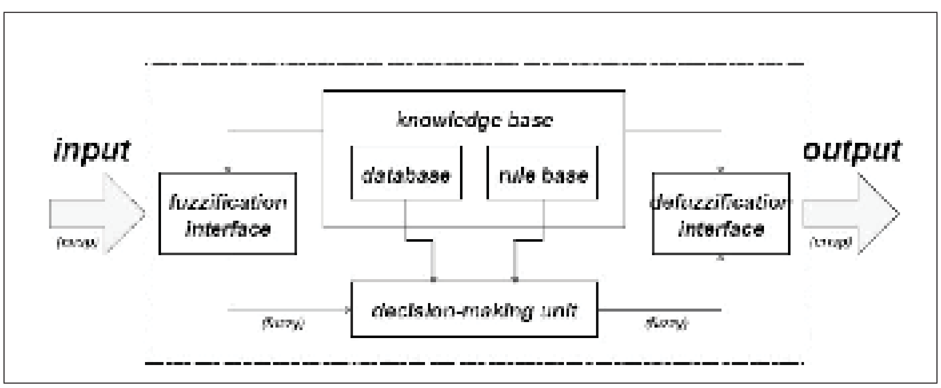

Figure 1.5 Architecture of ANFIS (Solatian *et al.* (2012))

Support Vector Machine (SVM) is a supervised learning algorithm that tries to maximize the margin between two class samples with finding some support vectors. If the samples are not separable in 2D space, it constructs a hyperplane to classify samples. It employs different kernel functions in order to mapping features between spaces. Many researches in the field of wear monitoring implemented SVM as a powerful and reliable classification approach (Pandiyan *et al.* (2018)).

Baysian networks have generally less accuracy but more reliability than other methods. The aim of these systems is to extract hidden knowledge from experimental data in the form of causal relationships and probabilities as well as using previous knowledge. They comparatively, need large training data and can be used for solving inverse problems such as finding optimal cutting parameters (Abellan-Nebot & Subirón (2010)). This method is used mostly for surface roughness and wear monitoring.

Hidden Markov Method (HMM) is also implemented in this field. HMM represents the probability distribution over a sequence of events over time. Zhang *et al*. (Zhang *et al.* (2010)) employed a hybrid model of Hidden Markov and ANN for chatter monitoring. It claims that this method outperforms the baseline method in terms of recognition and accuracy. Mathematical models can also be used as well to predict the tool faults. Based on the complexity of the nature of fault and based on how the signal reflects the faults, usefulness of mathematical models can be evaluated.

In a study, Wang *et al*. employed SVM, HMM and Radius Basis Function (RBF) to conquer the complexity of the machining process and the uncertainty of the tool wear evolution. They also implemented stacking ensemble strategy to reflect the relationship between the outputs of these base classifiers and tool wear states. Titanium alloy milling experiment was carried for validation and force signals were captured. The results show that ensemble strategy performs better in both classification accuracy and stability (Wang *et al.* (2014)).

1.3.2 Deep Learning

Deep Learning refers to machine learning techniques with deep architectures and multiple layers which enable them to learn highly complex relationships. In the current era that sensors are widely available and actively producing high amounts of data, it is necessary to develop methods which are able to extract most information out of the big data. In standard neural network (NN) which consists of many simple, connected neurons, input neurons get activated through the data from sensors. Other neurons are activated through weighted connections from previously layers' neurons. Learning process of the NN is about finding weights for each neuron that will result in desired output from the NN. The learning process may require long causal chains of computational stages(Schmidhuber (2015)). Rina Decher (Dechter (1986)) was the first to introduce Deep Learning in machine learning in 1986 and Igor Aizenberg in 2000 introduced deep artificial neural networks. Deep Learning is about accurately assigning the weights across many such layers and deep architecture, which enables it to learn highly complex relationships from even low-processed to raw signals (Wang *et al.* (2016)).

Training of deep NNs with many layers, had been difficult in practice by the late 1980s due to its high computational requirements. Recently, with the improvement of computing power and introduction of GPU, deep NNs have finally attracted widespread attention. Deep Learning methods are able to make the most information out of the big data and have powerful characteristics to outperform other methods when the relationship between the input data and desired output are complex (Pan & Yang (2009)). Convolutional neural networks (CNNs) and Recurrent neural networks (RNNs) are the most popular Deep Learning models.

CNNs were proposed by LeCun (LeCun *et al.* (1990)) for image processing, featured spatially shared weights and spatial pooling. CNN models have shown their success in various computer vision applications (LeCun *et al.* (1990), Jarrett *et al.* (2009), Krizhevsky *et al.* (2012)) and sequential data including Natural Language Processing and Speech Recognition (Abdel-Hamid *et al.* (2012), Kim (2014)). CNN uses convolutional layers and pooling layers to extract and learn abstract features from the input data. The convolutional layers (convolutional kernels) generate translation-invariant local features by convolving multiple local filters with raw input data and the subsequent pooling layers use sliding windows of the raw input data to extract features with a fixed-length (Zhao *et al.* (2016)).

RNNs are the deepest neural networks, which can address memories of arbitrary-length sequences of input patterns. Unlike mlp which only maps input data to target data, RNN is able to keep the entire memory of previous inputs in its internal state and map that to the target . RNNs can be trained with sequential input data and target outputs via backpropagation through time for supervised tasks (Funahashi & Nakamura (1993)). RNNs' main drawback is the vanishing gradient problem during backpropagation for model training, which is caused by the fact that neurons are independent of each other's history therefore RNNs may not capture long-term dependencies from other neurons. Therefore, Long-Short Term Memory (LSTM) and Gated Recurrent Units (GRU) were presented to prevent backpropagated errors from vanishing or exploding (Zhao *et al.* (2016)).

Despite the high potential of deep NNs, they are relatively recent in the field of machinery condition monitoring. In a study, Jing *et al*. developed a CNN based algorithm for gearbox condition monitoring (Yan & Lee (2005)). Zhao *et al*. conducted a survey study to investigate the applications of Deep Learning methods in machine health monitoring (Zhao *et al.* (2016)). In another study, Zhao *et al*. employed a modified version of LSTM networks (CBLSTM) for tool condition monitoring in milling process (Vincent *et al.* (2008)). In another study, vibration signals of a gearbox system are preprocessed using statistical measures from the time domain and frequency band energy from frequency domain. Afterwards, the feature vector is fed to CNN to train it to detect gearbox faults (Salakhutdinov & Hinton (2009)). Based on the literature review, there is a huge potential to apply Deep Learning methodologies in the field of tool condition monitoring.

1.3.3 Transfer Learning

Despite the advantages of machine learning and Deep Learning algorithms, they come with certain disadvantages. First, they require a lot of data for training phase which means high costs for data acquisition and initial expense to design a monitoring model. Moreover, machine learning models work under the conditions they are trained on an may not work well or be relevant on a similar but different conditions. These drawbacks can be addressed using a new technique in machine learning called transfer learning. It refers to the use of knowledge learned in one domain and applying it to a different but related domain (Pan & Yang (2009)). It is highly beneficial especially in industrial applications which recording data for training the model is challenging and expensive. Transfer learning is about leveraging the existing knowledge from the labeled data and avoiding excessive efforts for generating large labeled datasets for a similar task of interest and it is specially useful for the tasks with lower available data.

Transfer learning is widely used in image recognition and computer vision applications. Image classification, face recognition and similar models are trained using the large available datasets and new models can be developed by fine-tuning pre-trained Deep Learning models from large image dataset to use in other image recognition tasks. Muralidharan *et al*. leveraged transfer learning in medical imaging application by starting from an ImageNet pre-trained model. They reported that the main advantage of this method is when the amount of training data was limited (Muralidharan & Sugumaran (2012)). Natural language processing filed is also widely explored the transfer learning framework to gain knowledge from large available datasets and benefit it in another model. In a study, a bidirectional LSTM method is proposed for multiple tasks such as emotion classification and intensity regression on tweets data. A set of word2vec word embeddings is used which were trained on a large collection of 550 million Twitter messages. Afterwards, the researchers pre-trained the Bi-LSTMs on the dataset of Semeval 2017 and finally re-tuned it for similar new tasks (Bengio *et al.* (2013)).

Considering the capabilities of transfer learning to work with lower number of experiments and transfer knowledge between problems, it has a huge potential in the field of condition monitoring. In a study, Guo*et al.* proposed a deep convolutional transfer learning architecture in condition monitoring. Their methodology consists of two steps, condition recognition and domain adaptation. The first step is designed to automatically learn features and recognize health conditions of machines and the latter facilitates the first step to learn domain invariant features. Transfer learning approach is validated in this research using multiple datasets (Guo *et al.* (2018)). Literature review reveals the high potential of exploring transfer learning in the field of tool condition monitoring to address some of the issues in this field and leverage scalability between systems and models(Malhi & Gao (2004)).

1.4 State of the Art Summary

In summary, indirect monitoring methods are identified as one of the most practical solutions for tool condition monitoring. These approaches consist of signal acquisition, signal processing and noise reduction and modeling and decision making. This research will investigated force, current and vibration signals which are among the most common and practical signals for this application. On the signal processing side, time-frequency transformation is the advanced method which has the benefits of both time domain and frequency domain analysis and it became more practical recently due to the availability of higher computational resources. More specifically, spectral subtraction is identified as an advanced time-frequency processing techniques which is beneficial in the tool condition monitoring application.

In the decision making step, Deep Learning algorithms are state of the art with a lot of successful implementations in many industries. There is still a huge potential to explore different architectures and forms of these algorithms in various condition monitoring applications including tool condition monitoring. Transfer learning is also introduced as an emerging framework to solve some of the challenges in commercializing Deep Learning. This method can help reduce the high data requirements of Deep Learning algorithms and provide a framework to share knowledge between different models. In this research, we focused on the state of art methods to explore the presented approaches to enhance tool condition monitoring in milling operation.

CHAPTER 2

EXPERIMENTAL DESIGN

The first step of this project is to perform a set of experiments in order to obtain the relevant data from the machining process using sensors and data acquisition techniquse. First, input variables (factors) known as controlled and uncontrolled variables of the experiments are defined and explained, and then the needed devices and materials for performing the measurements are selected and finally, the measurement of tool wear is explained.

2.1 Inputs: controlled and uncontrolled variables

Controlled variables (factors) in an experiment are the variables that we have control over them and can choose the values for them in order to perform the experiments. In the following sections a list of these variables is presented and their specifications are presented.

2.1.1 Cutting Material

High Speed Tool Steels are good candidates in order to investigate tool wear in machining hard material with hardness of 60-62 HRC. D2 High Speed Tool Steel is selected as the cutting material (Figure 2.1). The material is cut to approximate dimension of 200 $\times$ 54 $\times$ 4 mm using a wet saw (Figure 2.2) to accommodate the size for the jig and machine clamp. Table 2.1 summarizes workpiece specifications.

Table 2.1 Workpiece specifications

Material	Steel D2
Hardness	60 -61 HRC
Workpiece dimensions	200 mm $\times$ 54 mm $\times$ 4 mm

D2 is one of the high carbon and high chromium cold work type of tool steels. This alloy is a deep hardening, highly wear resistant alloy. It is used for long run tooling applications where

Figure 2.1 D2 work piece

Figure 2.2 Work pieces are cut with wet saw

wear resistance is important, such as blanking or forming dies and thread rolling dies. The machinability of D2 is poor (D2). Table 2.2 presents the chemical composition of the material.

Table 2.2 D2 High Speed Steel Chemical Composition (D2)

Carbon	**Chromium**	**Molybdenum**	**Vanadium**	**Silicon**	**Manganese**
1.05%	12%	0.75%	0.80%	0.30%	0.45%

2.1.2 Cutting Tool

Based on recommendation of experts in the field of machining, the tool Walter End Mill Protostar H50 Ultra: AH8083128-1 with 6 teeth is selected as the cutting tool which is suitable for hard material machining. Figure 2.3 and Table2.3 presents geometrical specifications of tool.

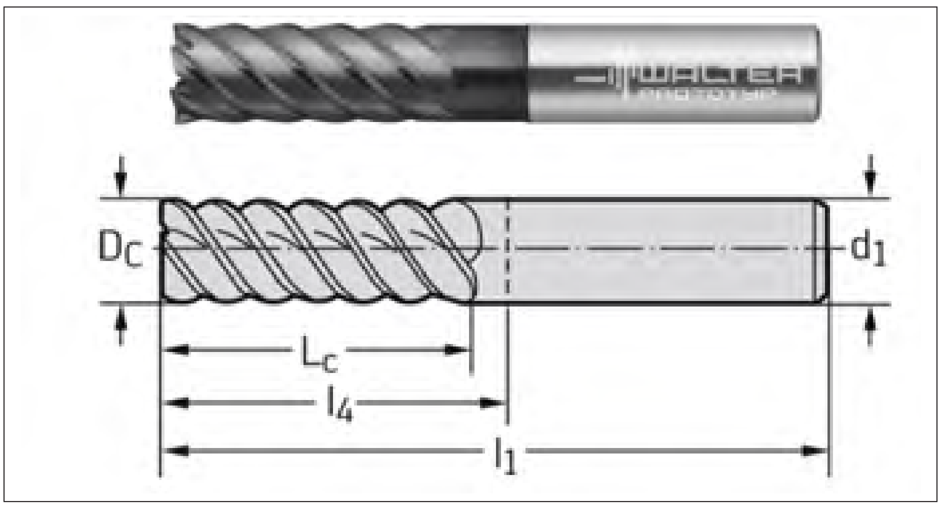

Figure 2.3 Tool geometry

2.1.3 Machine and operation type

K2X10 Huron CNC machine of the Laboratory LIPPS at ETS is chosen to perform the experiments. The machining centers, HURON K2X, enable machining operations in 3 axes, from roughing to finishing, of all kind of complex work pieces. Maximum power and torque specifications of the machine are presented in Table 2.4 With this machine, maximum length of workpiece must be 200 mm which has been taken to account in the workpiece dimensions and there should be a minimum of 3.5 mm in height for the clamping. Therefore, in designing the

Table 2.3 Tool specifications

Type of cooling	Dry
Tool	H8083128-12X26
Body material	Carbide
Helix angle	50 degree
Cutting edge diameter (Dc)	12 mm
Cutting edge diameter tolerance class	h10
Number cutting edges	6
Max. depth of cut (Lc)	26 mm
Usable length (l3)	26 mm
Overall length (l1)	83 mm
Maximum overhang (l4)	38 mm
Adaptor diameter, workpiece-side (d1)	12 mm

number of cuts in depth of cut direction we should take into account the amount needed for the clamping.

Table 2.4 Specifications of Huron K2X10 series

Machine available maximum power	25 KW
Machine available maximum torque	86 N.m
Maximum workpiece length	200 mm
Width of cut (a)	7.14 mm

Different types of milling are presented in Figure 2.4 The type of operation is considered to be side milling. Side milling (Figure 2.5) is preferred as it will not have different engagement at the wall in each cutting path. Tool engagement should be chosen considering the width of workpiece, in a way that maximum number of cutting paths is achieved. The preferred tool engagement is from 60% to 90% of the tool diameter as 100% engagement may cause burr around the edges of work piece and 50% causes high entry and exit shocks. Considering length of the workpiece which is 54 mm and diameter (ϕ) of 12 mm for the tool, we assume 7 cutting paths, with 60% of tool engagement. Width of cut will be a = 7.14 mm.To be able to perform the side milling with no engagement at the at the bottom of the tool a specific jig

(Figure 2.6) was designed which was used to hold the workpiece in place for clamping to the machining center.

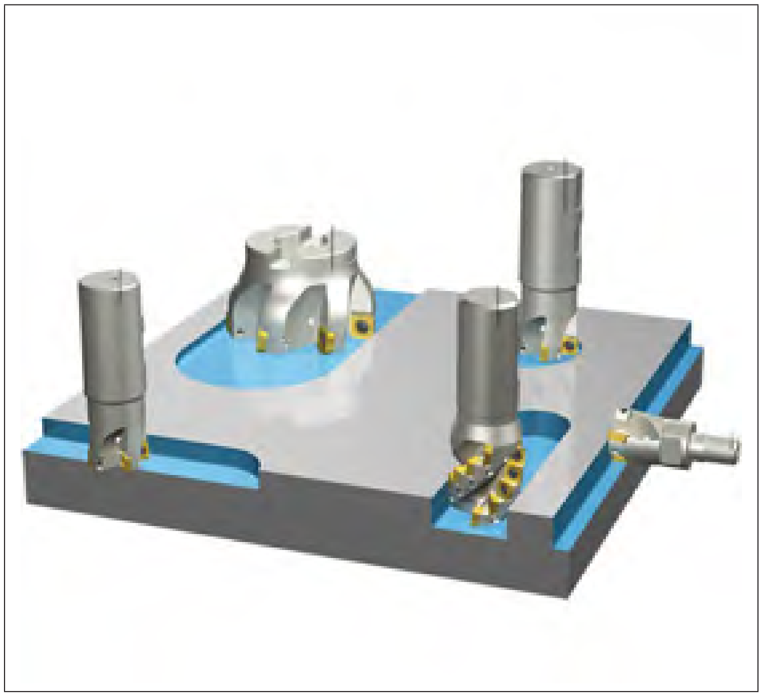

Figure 2.4 Different types of milling
(DirectIndustry (2016))

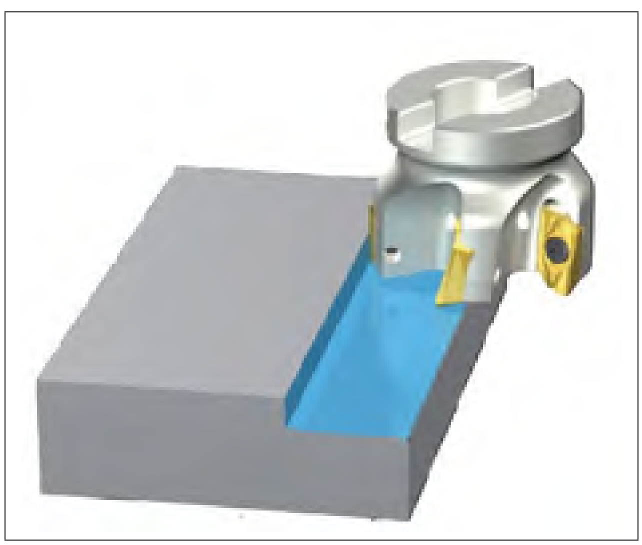

Figure 2.5 Side Milling (DirectIndustry (2016))

Figure 2.6   Jig designed for holding the work piece in place

2.1.4   Cutting parameters

Cutting parameters should be chosen with consideration of the tool, machine and workpiece specifications. These values should be chosen in a way that it does not exceed the maximum power and torque of the machine and also does not cause the tool to be damaged.

The first limitation in our case is the height of the workpiece that is 4 mm. We should choose appropriate amount of depth of cut in order to have enough number of cuts and also it should be a reasonable amount so that the tool can grab the material to cut. The tool manufacturer application is used to define the maximum and minimum range values that are possible to machine the specified material. During the experiments we encountered problem of tool breakage (Figure2.7) due to uneven surface of the workpiece while performing the first path causing higher engagement of the tool with the workpiece. to avoid this issue we need to make sure that the first path is performed with lower width of cut.

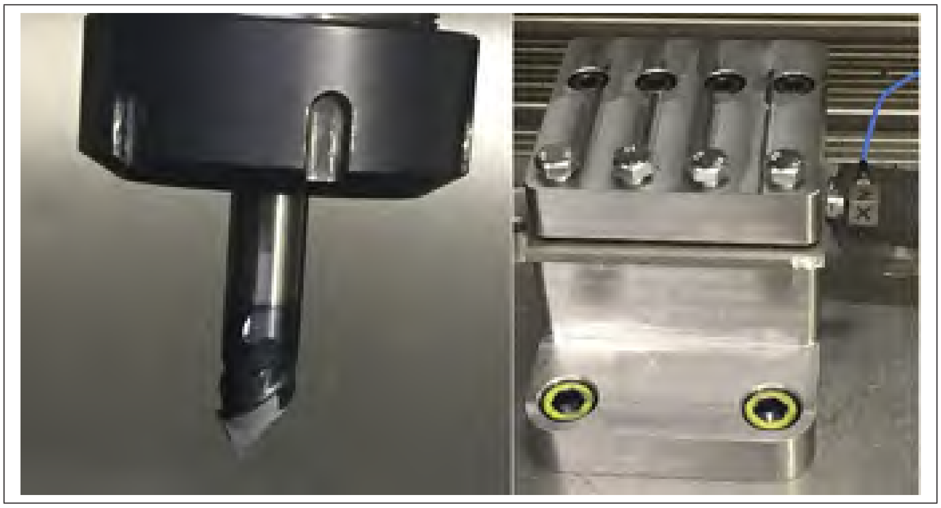

Figure 2.7 Tool breakage due to uneven surface of the workpiece in first path

2.1.5 Uncontrolled variables and noise

Uncontrolled variables are considered as inputs but we do not have control over them. These variables are mostly referred to as noise and can be considered in the following groups: the errors related to the devices such as measurement devices or machine itself: a range that is mentioned in the manual, the errors that their value is not known for us for example the non-homogeneity in the material which can cause different amounts of hardness in the material and finally, the uncertainty in tool wears measurement, which should be measured in the experiment.

2.2 Outputs and measurement sensors

Outputs are the variables that are going to be measured during the experimental testing. A methodology relates the outputs of sensors to tool wear. Table 2.5 is the list of outputs of the experiments and their measuring system.

Table 2.5 Sensors used for measurements

Outputs	Measuring device/sensor
Tool wear	Keyence VHX500 Digital Microscope
Force signal	Dynamometer Kistler® (9255B)
Acceleration signal	ICP Tri- axial Accelerometer (PCB356B21)
Electric current signal	Current sensor

After performing the machining in time intervals, tool wear measurements is performed with the VHX Keyence digital microscope (Figure 2.8) . Uniform flank wear (VB1) is measured as shown in figure 2.9 in all cutting edges of the tool and the average value is reported as the final tool wear. The measurement should be repeated several times for some cases to quantify uncertainty of the measurements.

2.3 Experiments framework

The purpose of experiments in this project is to create tool wear and perform measurements of vibration, force and power to relate the selected features of sensor signals to the measured amount of tool wear. Figure 2.10 shows the outline of the experiments including controlled and uncontrolled input parameters, modeling, process, and outputs of the experiments. The correlation between features from signals in time/frequency/ time-frequency domain and the amount of tool wear is the estimation model of tool wear, which will enable us to estimate the amount of tool wear with monitoring the process with sensors.

The experiments was conducted with two different cutting speeds of 2500 rpm and 6000 rpm and feed rates of 0.12 mm/tooth and 0.05 mm/tooth, with 4 mm depth of cut constructed multiple experiments. In our design of experiments we defined two speeds 2500 rev/min and 6000 rev/min with the maximum sampling frequency of 128 kHz. Two ICP tri-axial accelerometers with sensitivity of 10.08 mV/g and max frequency range of 10 kHz in Y and Z-axis and 7 kHz

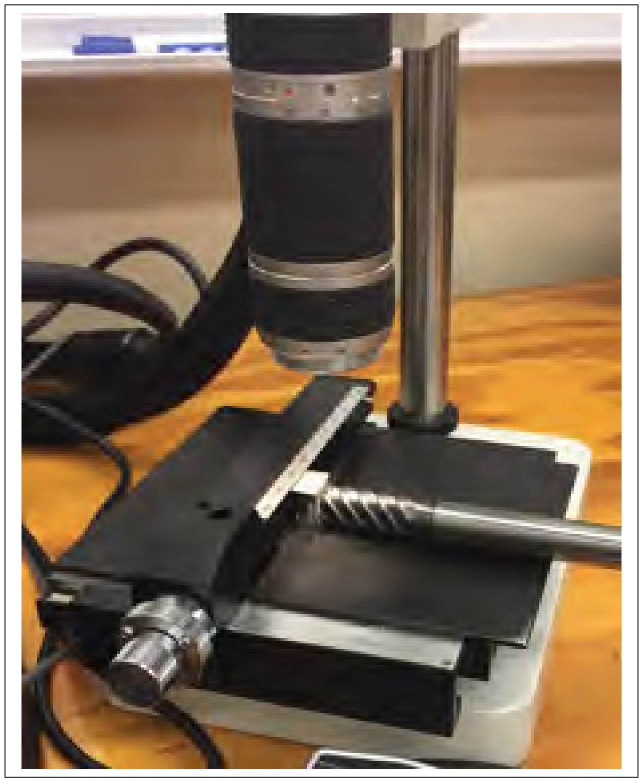

Figure 2.8 Keyence microscope used for tool wear measurement

in X-axis are used. They are placed at the spindle and table of the machine and then connected to the PCB ICP conditioners (Figure 2.12) and then to the data acquisition system DT9836 .

Force is also measured with the multi- component Kistler dynamometer 9255-B, table sensor with measurement range of -20 to +20 kN for x and y directions and -10 to 40 kN for z direction which is then coupled with charge amplifiers and then to the data acquisition system. Figure 2.11 presents sensor placements and data collection unit . The DT9836 board (Figure 2.13) used for sampling has a sampling rate for each channel of 225kHz and included an anti- aliasing filter for the force and vibration measurement. Aliasing is the phenomenon that frequencies greater than the Nyquist frequency are shifted erroneously to lower frequencies. According to the Nyquist sampling theorem, the sampling rate must be greater than twice the maximum frequency component of the signal of interest. In other words, the maximum frequency of the input signal must be smaller than half the sampling rate. For example, if the maximum

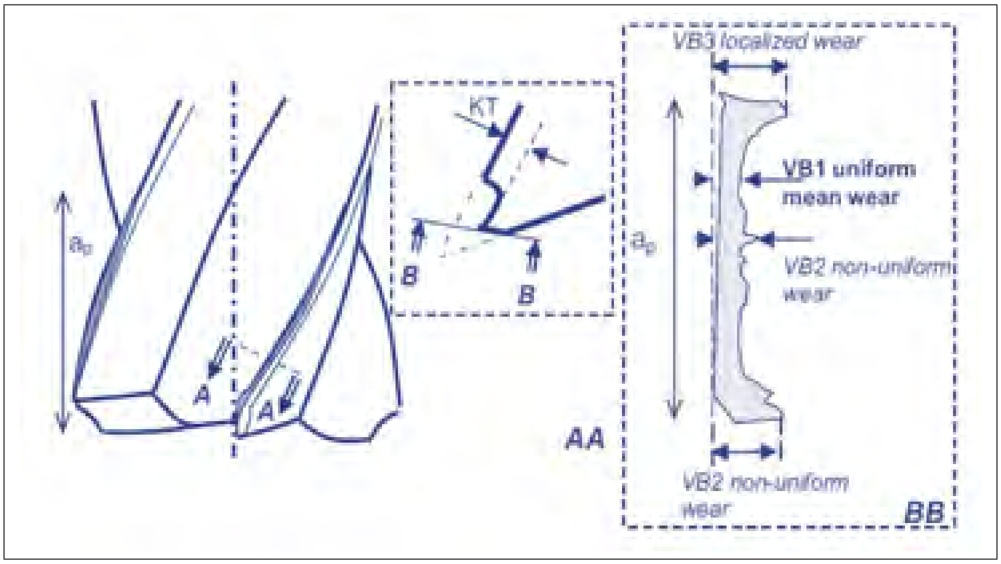

Figure 2.9 Tool wear measurement (ISO8688-2 (1989))

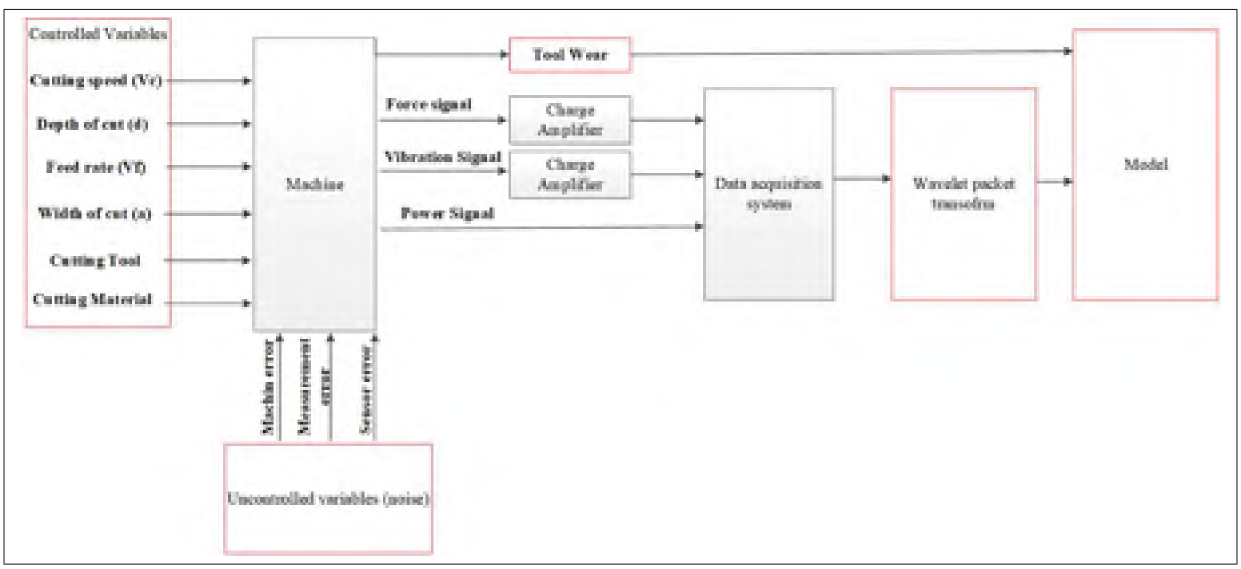

Figure 2.10 Outline of the experiments including input controlled and uncontrolled parameters, modeling, process, and outputs

frequency component of a signal is 1K Hz, the sampling rate must be greater than 2K Hz. In real-world applications, you can set the sampling rate between 3K and 5K Hz.To ensure that you limit the frequency content of the input signal, you can add a lowpass filter before the sampler and the analog to digital converter (ADC). A lowpass filter passes low frequencies

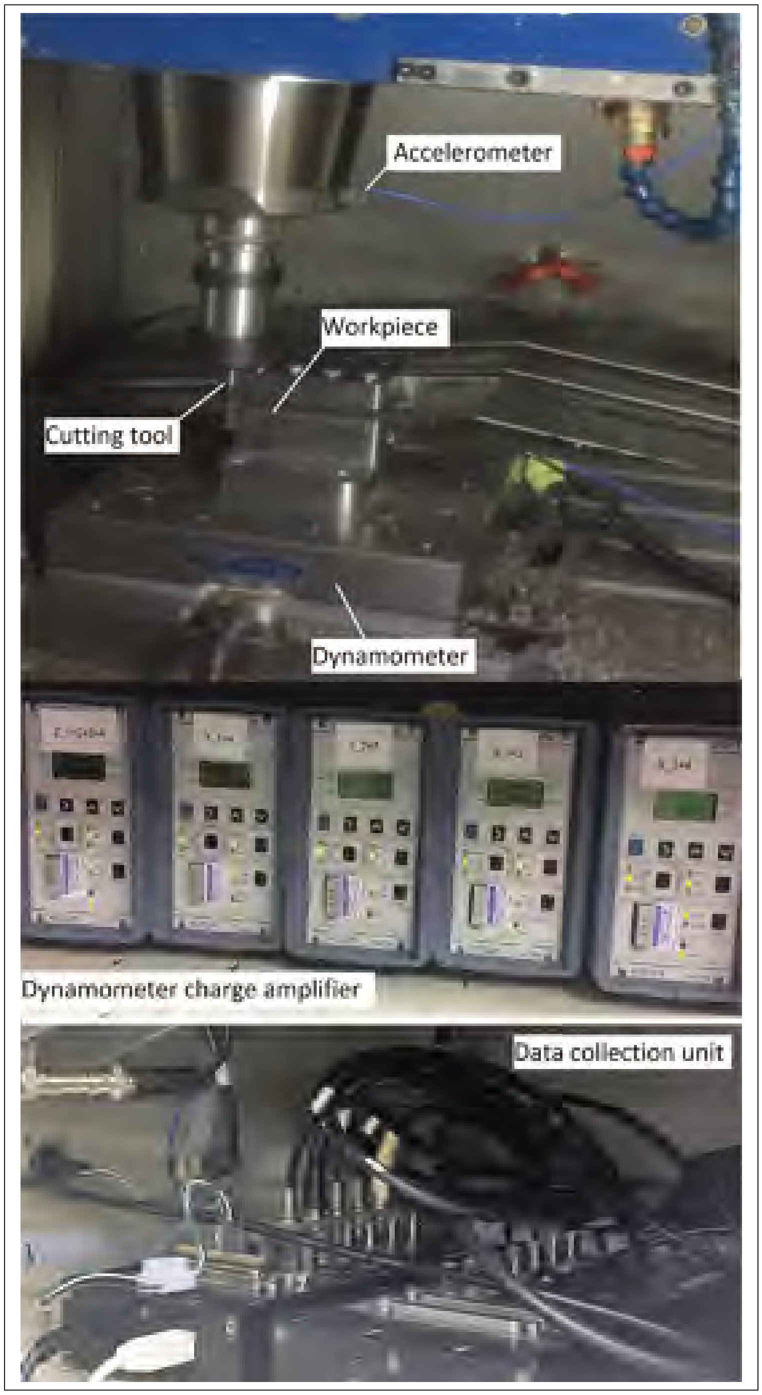

Figure 2.11 Sensor placements and experimental data collection

and attenuates high frequencies. This filter is an anti-aliasing filter because by attenuating the frequencies greater than the Nyquist frequency, the filter prevents the sampling of aliased

Figure 2.12 ICP conditioner for accelerometers

components. When you use a filter before the sampler and ADC, the anti-aliasing filter is an analog filter with a proper cut-off frequency. The cut-off frequency equals the maximum frequency component of the signal of interest. Using the anti-aliasing filter satisfies the Nyquist sampling theorem. Usually aliasing protection is automatic in any acquisition system. The only way to protect data from aliasing is to apply appropriate aliasing protection before the data is generated or acquired. Aliasing occurs when the data is generated or sampled. You cannot remove aliased components from the data without detailed knowledge of the original signal. In general, you cannot distinguish between true frequency components and aliased frequency components. Therefore, accurate frequency measurements require adequate aliasing protection.

LabVIEW software is used to data acquisition and recording. Tool wear were measured at different intervals which results in 63 cases with different tool wears and cutting conditions. Figure 2.14 represents a sample of force signal in LABVIEW software and Figure 2.15 shows a sample of vibration signal plotted by Python software.

Figure 2.13 Data collection
unit -DT9836

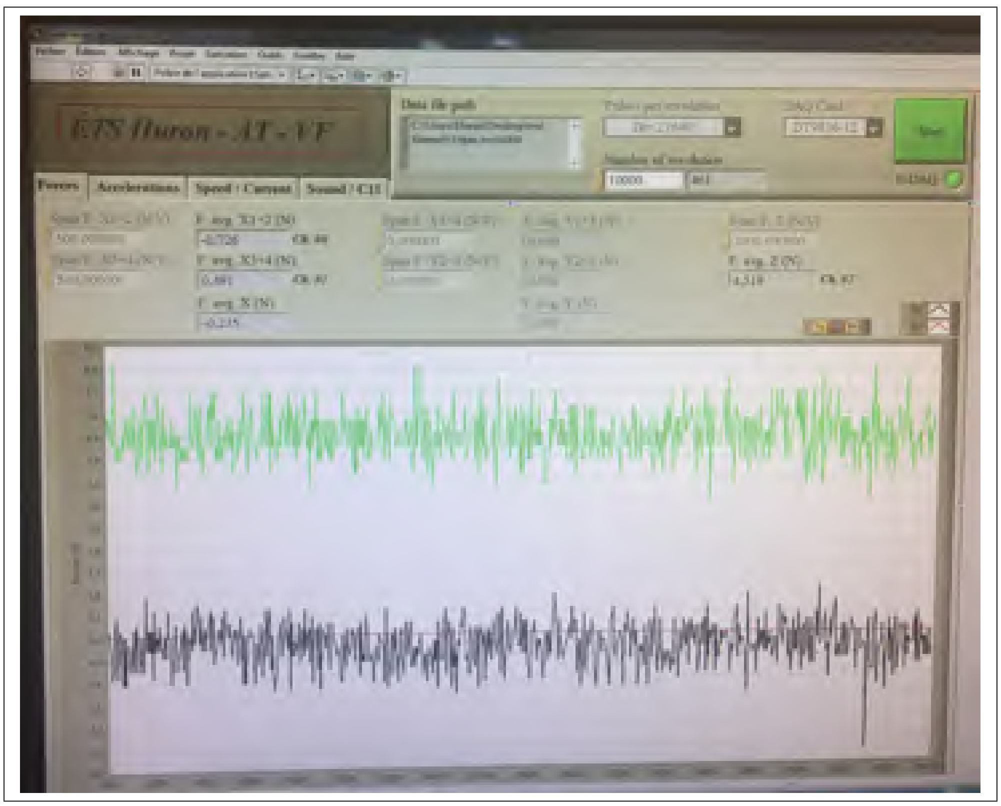

Figure 2.14 Data acquisition - Force signal

After data acquisition the signals are transformed to time-frequency domain using wavelet transform and wavelet packet transform from PyWavelets (Wasilewski (2010)). Wavelet packet

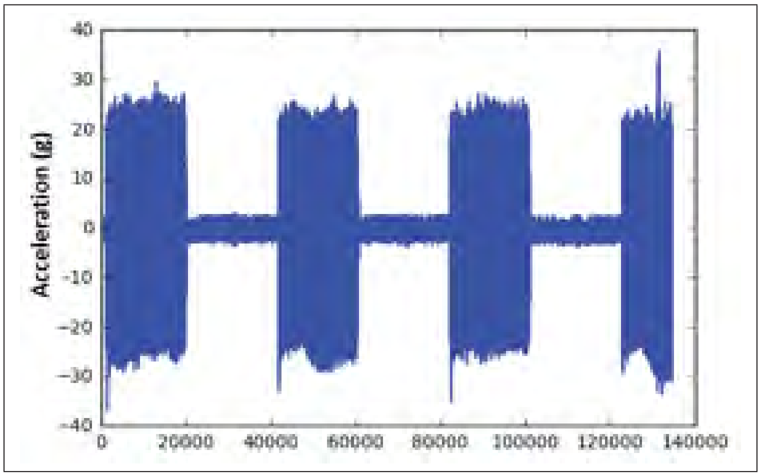

Figure 2.15   Data acquisition - Acceleration signal

transform is employed as it permits decomposing signals into uniform frequency bands. The wavelet packet transform method decomposes a time signal into several independent time-frequency signals called packets. Using the WPT, we can determine a signal's time–frequency composition, thereby having a good understanding of what is contained within the signal. Furthermore, the WPT can be applied to remove noise contained in the signal. Wavelet packet transform results in equal-width subband filtering of signals as opposed to the coarser octave band filtering found in the DWT.Note that with this filtering with wavelet decomposition for components with frequencies higher than frequency range of th sensors will be null components. In mathematics, the Morlet wavelet (or Gabor wavelet) is a wavelet composed of a complex exponential (carrier) multiplied by a Gaussian window (envelope).For n levels of decomposition the WPD produces 2n different sets of coefficients (or nodes) as opposed to (3n + 1) sets for the DWT. However, due to the down sampling process the overall number of coefficients is still the same and there is no redundancy. While discrete wavelet transform provides flexible time–frequency resolution, it suffers from a relatively low resolution in the high-frequency region. This deficiency leads to difficulty in differentiating high-frequency transient components. The wavelet packet transform (WPT), in comparison, further decomposes the detailed information of the signal in the high-frequency region, thereby overcoming this limitation. Figure 2.16 schematically illustrates a WPT-based signal decomposition process, where a four-level WPT produces a total of 16 subbands, with each subband covering one-sixteenth of the signal frequency spectrum. The analytical procedure of the DWT method is composed of low-pass and high-pass filters, respectively. Low-pass filters remove high frequency fluctuations and preserve slow trends to obtain an approximate signal. High-pass filters remove the slow trends and preserve high frequency, providing detail information. The output of low-pass filter sand high-pass filters provide the approximation and detail coefficients, respectively. This procedure is repeated until the desired wavelet decomposition level is achieved. The enhanced signal decomposition capability makes WPT an attractive tool for detecting and differentiating transient elements with high-frequency characteristics (Barros & Diego (2006)).

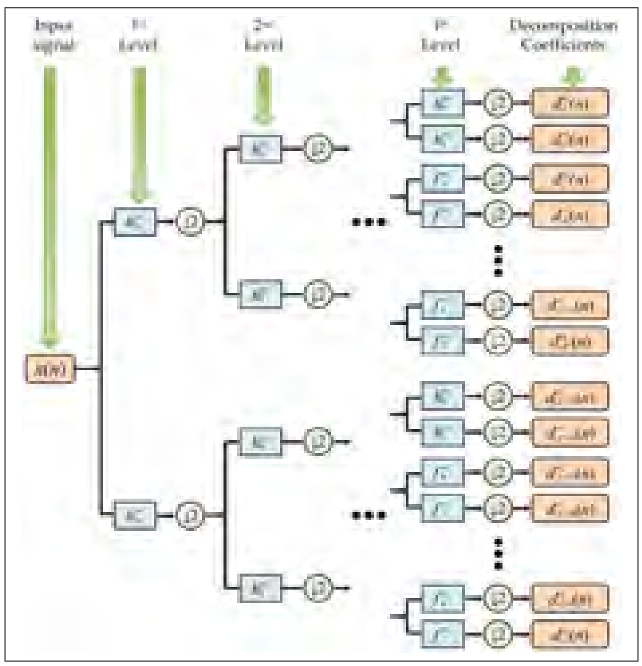

Figure 2.16 Wavelet packet transform

WaveletPacket function from PyWavelets (Wasilewski (2010)), the scientific Python module for Wavelet Transform calculations is used to perform the transformation. Morlet is used as the wavelet function and with 4 as the levels of decomposition and the output of the transformation is divided into 16 uniform bands which is used as the inputs of the models. Depending on th each analysis and model, further processing may be applied for the signals which each is presented in the corresponding chapters.

CHAPTER 3

TOOL CONDITION MONITORING USING SPECTRAL SUBTRACTION METHOD AND ARTIFICIAL INTELLIGENCE METHODS IN MILLING PROCESS

Fatemeh Aghazadeh[1], Antoine Tahan[1], Marc Thomas[1]

[1] Département de Génie Mécanique, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Published in International Journal of Mechanical Engineering and Robotics Research, January 2018

Abstract

Process monitoring is necessary in machining operation to increase productivity, improve surface quality and reduce unscheduled downtime. Tool wear and breakage are important and common sources of machining problems due to high temperature and forces of machining process. Therefore, it is highly beneficial to develop an online tool condition monitoring system. This paper investigates a robust tool wear monitoring system for milling operation. Spindle current is employed as the fault indicator due to its cost-effectiveness and ease of use in industrial environment. Wavelet time-frequency transform is used a superior tool to simultaneously investigate time-varying characteristics of the signal and its frequency components. After the time-frequency step, spectral subtraction algorithm is employed to intensify the effect of tool wear in the signal and reduce the effect of other cutting parameters. Based on this method, the average signal spectrum of healthy case is subtracted from all the signals with the same cutting parameters. After further processing and noise reduction, fault features and indicators are extracted from the results of the processed signal. Finally, five advanced machine learning algorithms are implemented for modeling the system. Gaussian process regression, support vector regression, Bayesian rigid regression, Nearest neighbor regression and decision tree methods are compared.The methods are validated based on the experimental data.Results show high accuracy for the tool wear estimation while decision tree method was superior to others with accuracy of 91.58%.

3.1 Introduction

Machining processes, especially milling operation is fundamental in today's manufacturing industries. There is a growing demand to make the machining operation totally automatic to increase the productivity. Along other directions in automation, it is necessary to automatically monitor machining online to assure the safety and part quality. Tool defects can be considered as one of the most common and costly faults of the machining process. Due to the contact forces and friction between cutting tool and workpiece, high temperature in the cutting area and pressure of the chips on the tool, some defects may happen to the tool which deteriorates the surface finish or cause damage or breakage to the tool, workpiece or machining center (Zhu & Vogel-Heuser (2014)). Therefore, it is in high demand to design a reliable and robust online automatic TCM system to improve the accuracy, reduce the production cost and increase the productivity.

TCM methods can be categorized into two main groups: direct and indirect methods. Direct methods use directly measured actual value of fault with sensors such as laser, optical and ultra-sonic sensors. However, in indirect methods, physical parameters of the system such as force, vibration etc. are utilized to represent tool condition, indirectly (Siddhpura & Paurobally (2013)). Although direct measurement methods estimate tool fault with high accuracy, they are still expensive and not suitable for online application in industrial environment. However, indirect methods can be used to fulfill TCM purpose as an alternative with accurate results and acceptable cost by using a proper descriptor signal and an appropriate modeling method (Abellan-Nebot & Subirón (2010)).

Signals which are most widely used for tool condition monitoring includes: Force, vibration, acoustic emission, current and power signals. Wang *et al.* studied tool wear monitoring using force signal and a distributed Gaussian ARTMAP (DGAM) network (Wang *et al.* (2013)). While force signal shows promising behavior to represent tool wear variations during the machining process, it is also highly dependent to other operating conditions and relatively expensive to be used in industry. In another study, acoustic emission signal is utilized for tool wear monitoring using a time-frequency signal processing approach (Rad *et al.* (2014b)). Vibration sensors are also very practical to be used in industrial environment. For example, in an study by Gangadhar *et al*. condition of a single point cutting tool is monitored with help of the vibration signals acquired from an accelerator (Gangadhar *et al.* (2014)). Soltani Rad *et al*. also employed current of the spindle motor sensors as an economic and practical indicator signal and achieved acceptable results for monitoring of the tool flank wear condition (Rad *et al.* (2013)). To make monitoring systems more robust, sensor fusion is an powerful approach. Sensor fusion refers to combining the information of more than one sensor in a complementary way. For example, Cho *et al*. combined information of force, vibration, acoustic emission, and spindle power sensor in time and frequency domain to develop an accurate and robust monitoring system (Stockwell *et al.* (1996)).

After choosing appropriate sensors and signal acquisition, the signals should be processed to reveal the effect of monitoring variables and remove noises. Time domain analysis, frequency domain analysis and time-frequency domain analysis are three common approaches for signal processing stage (Rehorn *et al.* (2005)). While many researches are devoted to time domain and frequency domain analysis, there are not many studies on time-frequency analysis in this area. Based on the non-stationary nature of faulty signals, time-frequency analysis can provide discriminative information about machinery health conditions. Therefore, discriminative fault features can be extracted from a faulty signal by choosing a proper time-frequency method as it considers frequency domain and time domain information at the same time (Feng *et al.* (2013)). In a study, Rehorn*et al*. proposed a feature extraction method in time-frequency domain called selective regional correlation for machining faults monitoring (Rehorn *et al.* (2006)). In another study, five different time-frequency transformation methods are employed and compared for the purpose of TCM (Rad *et al.* (2014a)).

The output of time-frequency domain has high dimensions. Therefore, after time-frequency analysis, this information should be converted to the appropriate feature vectors to make the monitoring problem solvable. Dimensionality reduction methods such as PCA and LDA are popular among the literature to perform this task. Spectral subtraction is another method which can be implemented to enhance the signal quality. It it originally introduced in the speech enhancement and recognition field to remove the effect of steady sounds in the environment (Cho *et al.* (2010)). While in the sound and speech analysis domains it is applied as a noise reduction method, in fault diagnosis applications it can be employed to present fault indicators. For example, El Bouchikhi *et al*. Proposed an algorithm for fault diagnosis of induction machine bearings using Spectral subtraction method. In this study stator current frequency response of the healthy machine is subtracted from spectrum of machine current acquired signal to present better fault indicators (Zhao *et al.* (2017)). Jing *et al*. developed a CNN based method for gearbox condition monitoring using frequency data of vibration signals and their method outperformed some of the common machine learning algorithms. Based on high potentials of these method, further researches are crucial to examine them with different signals and levels of signal processing in tool condition monitoring application.

During the machining process, many parameters such as operation conditions, depth of cut, feed rate and workpiece material changes which may degrade the monitoring system performance and can be a reduce system robustness. Therefor, a model between the prepared feature vectors and tool condition should be developed with ability to represent non-linear complex systems. Many methods such as artificial neural networks, Fuzzy logic, Neuro-fuzzy , support vector machine (SVM) and Bayesian networks are employed to perform this task in the literature. While these methods are individually are implemented and used in the literature, a comparative study between them could be beneficial for the researchers in this domain.

In this study, a TCM system is developed based on current signal of spindle motor as the fault indicator signal. Wavelet time-frequency analysis method is employed for signal processing step. After the wavelet analysis, spectral subtraction method is applied around tooth path frequency. Based on this algorithm, an estimation of signal spectral response is calculated for healthy state and the result is subtracted from each new signal with the same cutting parameters. Further noise processing is performed on the output of spectral subtraction step and a number of features are generated to represent the fault in the signals. Gaussian process regression, support vector regression, Bayesian rigid regression, Nearest neighbor regression and decision tree methods are implemented to learn a model between tool wear behavior and signal indicators. This paper is organized as follows: Section 2 represents the methodology and algorithm of monitoring system. Backgrounds and formulation of the methods which are used in this paper are explained in section 3. Section 4 introduces the benchmark dataset for validation of this work. Results and discussion are presented in section 5 and section 6 is dedicated to conclusion.

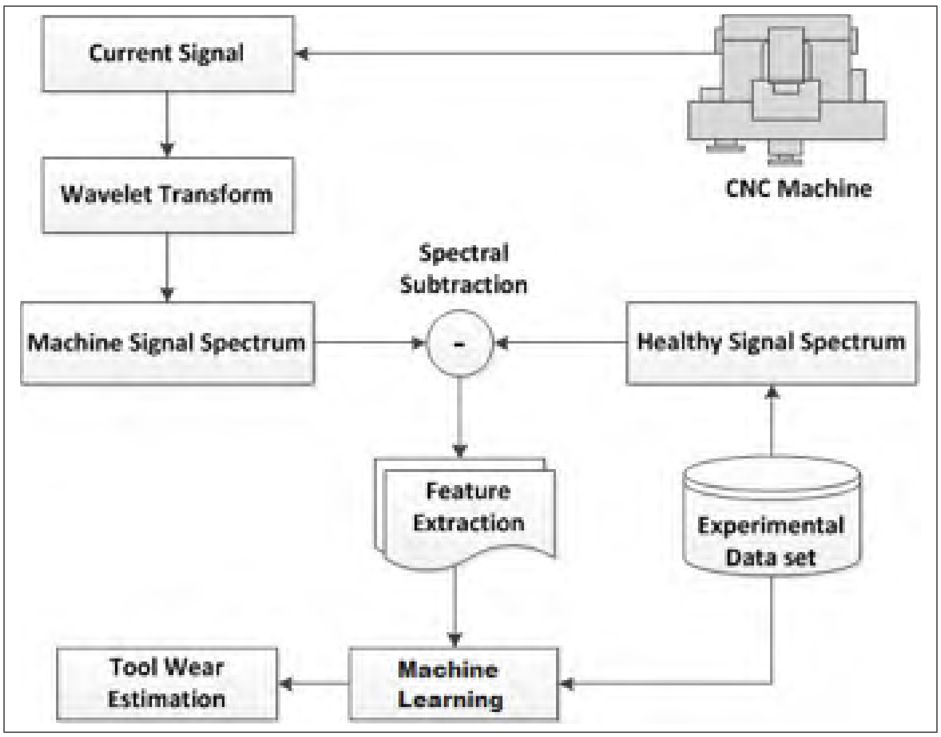

Figure 3.1 Methodology diagram

3.2 Methodology

This study investigates Tool wear Monitoring using time-frequency transformation, spectral subtraction and machine learning. Figure 3.1 depicts the monitoring system's methodology diagram. The fault descriptor of this research is current signal based on its high performance and applicability in industrial environment.After acquisition of signal, it will be transformed to time-frequency domain. Using the experimental data an estimate of spectrum of healthy tool for different cutting conditions is obtained. Afterwards, it is subtracted from each new signal under same cutting conditions. It helps to intensify the effect of fault and remove the steady state part of spectrum for normal situation. Finally a machine learning method is used to model the system using experimental dataset.

3.3 Background of methods

This section presents the formulation and background of the algorithms and techniques used in this paper. First, Wavelet Transform is introduced and its formulation is provided. Then backgrounds of spectral subtraction is presented.

3.3.1 Wavelet Transform (WT)

Wavelet transform is one of the methods that widely used for health condition monitoring systems in the literature. In wavelet transform, wavelets are used as the basis instead of sinusoidal functions that are used in Fast Fourier transforms. It is an effective tool for transient signal analysis as well as time-frequency localization since, it adds a scale variable in addition to the time variable in the inner product transform. It has a better time localization but a lower frequency resolution for higher frequency components. In contrast, for lower frequency components, the frequency resolution is higher while the time localization is worse. Following equation describes the formulation of the continuous wavelet transform (Feng *et al.* (2013)).

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\, \psi \frac{(u-t)}{a}\, du \tag{3.1}$$

where wavelet $\psi(u-t)/a$ is derived by dilating and translating the wavelet basis $\psi(t)$,and $1/\sqrt{a}$ is a normalization factor to maintain energy conservation and $a > 0$.

3.3.2 Spectral Subtraction

Spectral subtraction is a method which originally was used for speech signal enhancement. A signal is considered a combination of noise and clean speech, therefore the noise spectrum is estimated during speech pauses, and an estimation of the noise spectrum is subtracted from the noisy speech spectrum to obtain the clean speech. It can be employed in fault diagnosis applications by removing the steady state and normal process spectrum from the new signals to obtain their anomalies and fault signatures. Consider a measured signal which consists of the steady state normal component and additive fault (Feng *et al.* (2013)):

$$y[n] = s[n] + d[n] \tag{3.2}$$

where y[n],s[n] and d[n] are the sampled measured signal, fault and steady state component, respectively. The frequency domain representation of the signal is given by:

$$Y(jw) = S(jw) + D(jw) \tag{3.3}$$

Therefore, the fault component of the signal can be obtained based on the following equation:

$$\widehat{S}(jw) = Y(jw) - \widehat{D}(jw) \tag{3.4}$$

where $\widehat{S}(jw)$ is the fault related spectrum estimate and $\widehat{D}(jw)$ is estimate of steady state component of spectrum. $\widehat{D}(jw)$ often is obtained using the time-averaged signal spectrum using the normal healthy state of the system:

$$\widehat{D}(jw) \cong |\overline{D}(jw)| = \frac{1}{K} \sum_{i=0}^{K-1} |D_i(jw)| \tag{3.5}$$

3.4 Experimental dataset

Tool fault detection model and validation of the method of this research is implemented by the benchmark NASA Ames and UC Berkeley milling dataset (Agogino & Goebel (2007)). The experiments are performed under various operating conditions using the Matsuura MC-510V machining center. In this research, current sensor of spindle is selected based on its ease of use and practicality in industrial application. The dataset signals include cases with changes in depth of cut, feed rate and therefore, effect of these parameters can be investigated on the monitoring system accuracy and system will be developed under varying cutting parameters.

The tool is a 70mm face mill with 6 KC710 inserts based on its industrial applicability. Workpiece material in the research is cast iron. A OMRON K3TB-A1015 current converter feeds the signal from one spindle motor current phase into the cable connector and a model CTA 213 current sensor(Flexcore Div. of Marlan & Associates, Inc.) is used for data acquisition. Flank wear (VB), which is defined as the distance from the cutting edge to the end of the abrasive wear on the flank face is considered as the fault and its value is reported in all the experiments using a microscope.

3.5 Results and discussion

After signal acquisition, signals should be processed to extract better fault indicators and remove noises. Signals are transformed to time-frequency domain as the first step of the processing using Morlet Wavelet transform method. Figure 3.2 represents the result of wavelet transform. The diagram represents the WT output for the healthy signal (VB=0) as well as four states of the fault. It can be observed from the diagrams that there is high magnitudes around tooth pass frequency. As fault value increases, the magnitude and density of wavelet values are increased, However the data is still noisy and need to be further processed to extract discriminative features.

After wavelet analysis, spectral subtraction is applied to the signals. For this purpose, an average estimation of spectrum of current signal for healthy case is extracted from the dataset. For

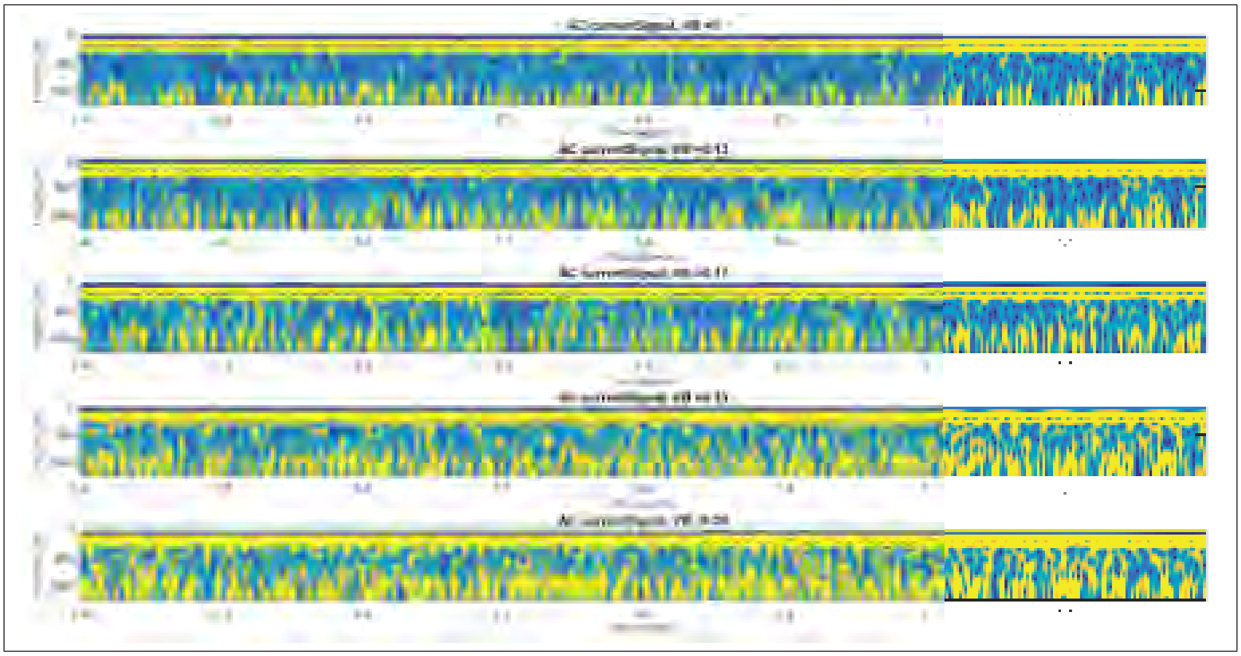

Figure 3.2   Wavelet representation of the signals with different VB levels.

each new signal of machine, the estimated healthy spectrum under the same cutting conditions is subtracted from signal. It can help to normalize the signals based on their cutting conditions and magnify the effect of tool wear by removing steady state components of the signal in normal situation. Figure 3.3 presents the result of spectral subtraction for different states of the fault. Based on the diagrams, for the healthy state, signal has zero or low magnitudes for most of the regions, as fault develops, the magnitude of spectrum specially around tooth passing frequency increases. The effect of fault and its progress is more clear in this graphs compared to previous representation (Figure 3.2).

Further noise canceling and signal refinement is performed after spectral subtraction step. for this purpose a local region around tooth pass frequency of signal is selected as the fault is most visible in this local window for further analysis. Moreover, the spectrum magnitude of the coordinates which are relatively less than (< 1/3 *spectral average) average magnitude of spectrum are set to 0. Figure 3.4 depicts the signals after further processing. As it can be seen from the graphs, tool wear signature is clear in the signals and therefore signals are ready for feature extraction step.

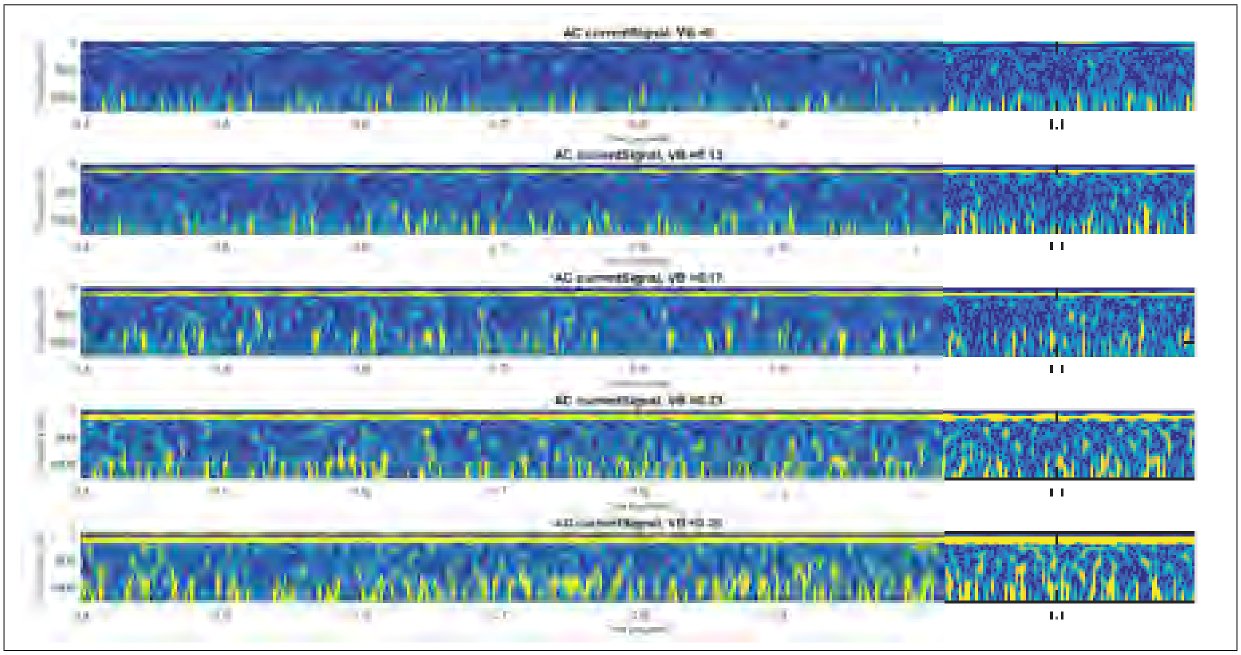

Figure 3.3 Wavelet representation of signals after spectral subtraction for different levels of VB

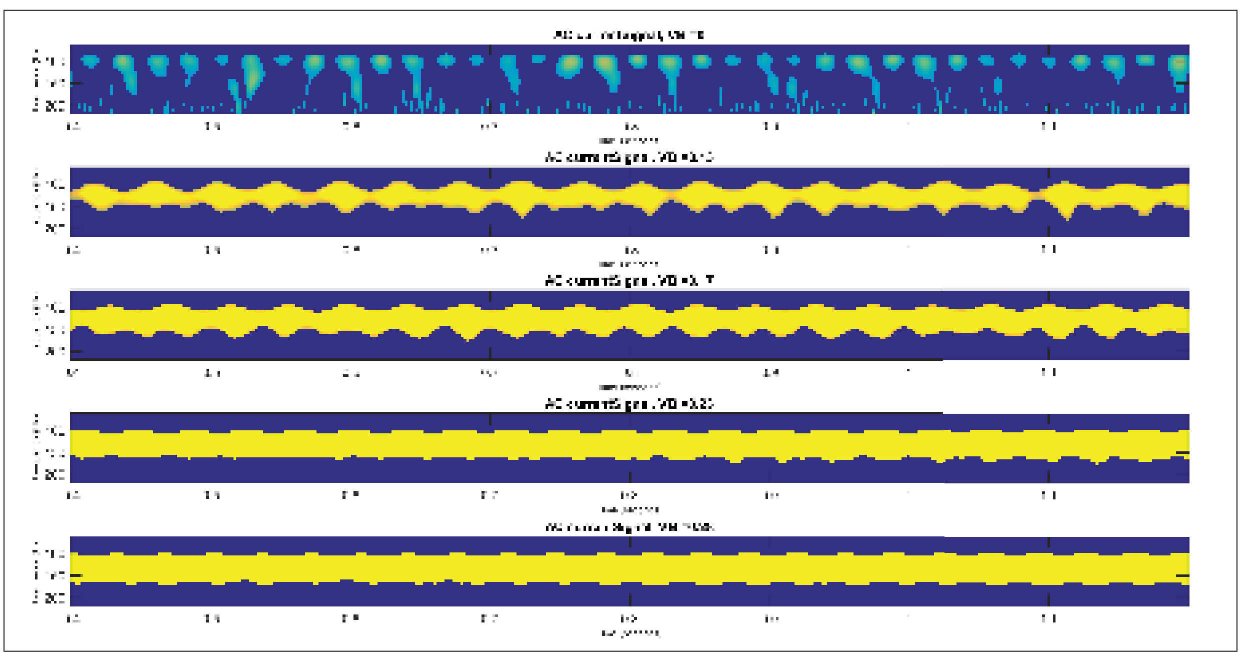

Figure 3.4 Signal representation after noise reduction for different VB values.

In the next step, various features are extracted from each signal. Lower band of signal, upper band of signal, maximum magnitude and frequency of its occurrence, variance, standard deviation and width of frequency response are among extracted features. Figure 3.5 shows lower and upper band of signal and width of frequency response.

The extracted features can change due to tool wear as well as other process characteristics. Changes in the cutting parameters such as depth of cut, feed rate and workpiece material may affect the features and therefore, makes defining a model for the machining monitoring more challenging. Generally, having more variables makes the system more complex to model. Therefore an model with ability to learn multidimensional non-linear relationships is necessary for the next step.

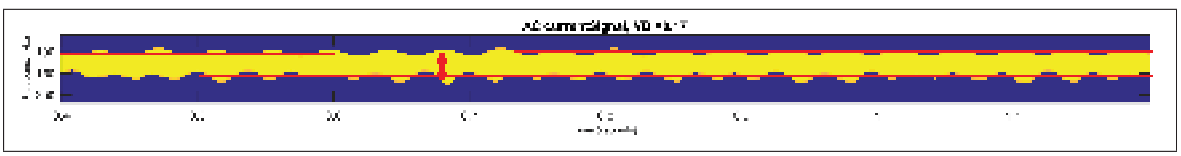

Figure 3.5 Comparison between TCM systems with changing in depth of cut, feed rate and workpiece material

Various machine learning and regression methods are employed as the last step for modeling the system. Experimental data is used to construct a dataset of feature vectors with their corresponding fault values. the monitoring systems are trained and tested under various scenarios. 80% of dataset samples are devoted to training step, and 20% are reserved for testing. For the monitoring system, average accuracy in percentage and RMSE are calculated as a representative of the performance. Table 1 presents results of the experiments for the system using test dataset. The systems are trained and tested under varying cutting parameters and fault indicators is current signal.

Based on the results, decision tree regression methods is superior to other methods with 91.58% accuracy in tool wear estimation. Bayesian rigid regression and nearest neighbor methods are also promising with 90.81% and 88.38% accuracy respectively. In general, all the methods provide acceptable accuracy (minimum for Gaussian process regression with 77.01% accuracy) which shows the robustness and high distinctiveness of the features in fault representation. Root mean square error is lowest for Baysian rigid regression (0.872) and almost similar to decision tree method RMSE (0.0877). Highest RMSE value belongs to Gaussian process regression with value of 0.2301. Figure 3.6 also depicts a comparison between the machine learning methods.

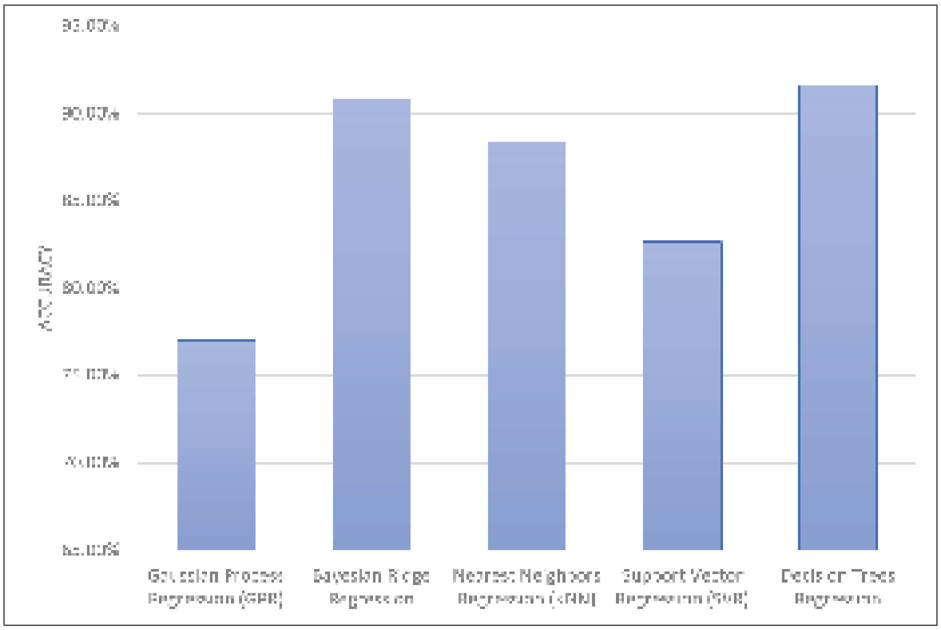

Figure 3.6 Results of Machine learning methods accuracy

Table 3.1 Comparison between accuracy and RMSE of different regression methods

Regression Algorithms	Average Accuracy %	Root Mean Square Error
Gaussian Process Regression (GPR)	77.01	0.2301
Bayesian Ridge Regression	90.81	0.0872
Nearest Neighbors Regression (KNN)	88.38	0.1131
Support Vector Regression (SVR)	82.61	0.1675
Decision Trees Regression	91.58	0.0877

3.6 Conclusion

In this research, tool condition monitoring under changing cutting parameters is investigated. A system is designed and developed for tool wear estimation which consists of: 1- Spindle current signal as a practical fault indicator, 2- an advanced time-frequency transformation method called Wavelet transform due to its great applicability to process signals and reveal rich information in both time and frequency domain simultaneously, 3- spectral subtraction method to remove the effect of normal operation from the signal and intensify the fault signature which helps to extract the most discriminative and relevant features to the fault and 4- Advanced machine learning methods to model the relations between the signals and their corresponding fault values. These methods are implemented to construct a model and estimate tool wear for new inputs based on the defined model. In overall, the algorithm proposed by this research showed accurate results with accuracy of up to 91.58% for tool condition monitoring with promising ability to tolerate and work under changing operation conditions.

Wavelet analysis revealed the time variant characteristics of frequency response of the signal and is beneficial in revealing fault characteristic of the signal. Therefore this study confirms its performance and applicability for this application. Spectral subtraction method also highly contributed in revealing the fault signature in the signal. This method removed the steady state part of the signal due to normal cutting and magnified remaining fault characteristics which proved its applicability and high performance in the processing step of the system.

Final step was a comparative study between state of the art machine learning methods for modeling the system. The robustness of the system and its performance using different methods are investigated. The results endorse the proposed methodology for wear estimation as all the systems has satisfactory accuracy for industrial application. Decision tree method has the highest accuracy of 91.58% for the test data set. Lowest RMSE belongs to Baysian rigid regression and decision tree methods. Highest RMSE is calculated for Gaussian process regression.

Acknowledgement

The authors gratefully acknowledge the experimental mill data provided by UC Berkely BEST lab and NASA Ames Prognostic Data Repository.

CHAPTER 4

TOOL CONDITION MONITORING USING SPECTRAL SUBTRACTION AND CONVOLUTIONAL NEURAL NETWORKS IN MILLING PROCESS

Fatemeh Aghazadeh[1], Antoine Tahan[1], Marc Thomas[1]

[1] Département de Génie Mécanique, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Published in INT J ADV MANUF TECH, July 2018

Abstract

Process monitoring is necessary in machining operation to increase productivity, improve surface quality and reduce unscheduled downtime. Tool wear and breakage are important and common source of machining problems due to high temperatures and forces of the machining process. Therefore, it is highly beneficial to develop an online tool condition monitoring (TCM) system.

This paper investigates a robust tool wear monitoring system for milling operation. Recent developments in machine learning, in particular deep learning methods result in significant improvement in automation of different industries. Therefore in this research we employed convolutional neural network (CNN) as a well-established and powerful deep learning algorithm for tool wear estimation. Wavelet packet based features are extracted for tool wear monitoring as a powerful time-frequency fault indicator. Moreover, a hybrid feature extraction method is proposed using Wavelet time-frequency transformation and spectral subtraction algorithms to intensify the effect of tool wear in the signal and reduce the effect of other cutting parameters. CNN based monitoring systems are compared with three other machine learning methods (Support Vector Machine, Bayesian Rigid Network and K Nearest neighbor method) as the baseline. The research is validated using different datasets. The algorithms are implemented and compared using experimental force and vibration signals from LIPPS lab of ETS university as well as using current signals as the fault indicator from Nasa_Ames dataset.

4.1 Introduction

Machining processes are fundamental part of today's competitive manufacturing industries. Due to the need for higher productivity, higher quality parts and cost reduction, there is growing demand to make the machining operation totally automatic. Along with other directions in automation, it is necessary to automatically monitor machining online to assure the production safety and quality. Tool defects can be considered one of the most common and costly faults of the machining process. Due to contact forces and friction between cutting tool and workpiece, high temperatures in the cutting area and pressure of the chips on the tool, various defects may happen to the tool which deteriorates the surface finish or causes damage or breakage to the tool, workpiece or machining centre (Zhu & Vogel-Heuser (2014)). Therefore, there is high demand to design a reliable and robust online automatic TCM system to actively monitor the cutting process and provides live reports of tool condition status.

TCM methods can be categorized into two main groups: direct and indirect methods. Direct methods directly measure actual value of faults with sensors such as laser, optical and ultrasonic. Another approach is used in indirect methods by employing physical parameters of the system such as force, vibration, etc. to represent tool condition indirectly (Siddhpura & Paurobally (2013)). Although direct measurement methods estimate tool fault with high accuracy, they are still expensive to implement and not suitable for online applications in industrial environments. However, indirect methods can be used to fulfill TCM purposes as an alternative with accurate results and acceptable cost by using a proper descriptor signal and an appropriate modeling method (Abellan-Nebot & Subirón (2010)). Moreover, the same sensor can be used for multiple monitoring tasks.

Applicable and informative signals that are widely used for TCM includes: force, vibration, acoustic emission, current and power signals. Li *et al*. studied TCM for turning process by employing force signals as the fault indicator (Li *et al.* (2017)). They extracted fourteen time-domain features from the force signals and using v-support vector regression, developed a model for flank wear estimation. While the force signal shows promising behavior to represent tool wear variations during the machining process, it is also highly dependent on other operating conditions and relatively expensive for industry use (Abellan-Nebot & Subirón (2010)). Vibration sensors are also practical in industrial environments. For example, Harun *et al.* studied TCM during deep twist drilling process and compared vibration and force signals for this purpose using time and frequency domain fault descriptors.Their study suggests that both sensors are capable of performing this task, but they recommended vibration signal as the superior fault indicator. (Harun *et al.* (2017)). Acoustic emission is also practical and informative signal which is highly used in the literature for TCM (Bhuiyan *et al.* (2016); Rad *et al.* (2014b)). Soltani Rad *et al.* also employed spindle current as an economic and practical indicator signal for tool breakage detection in milling process. In this paper the authors applied least squares support vector machine (LS-SVM) as the classifier and achieved acceptable results for monitoring tool breakage (Lin *et al.* (2017)). To make monitoring systems more robust, sensor fusion is a powerful approach. Sensor fusion refers to combining the information of more than one sensor in a complementary way to enhance the accuracy and reliability of the system. For example, Segreto *et al.* employed cutting force, acoustic emission and vibration signals for tool condition assessment in turning process. Signals are fused in feature level after processing and the results are fed to a neural network (Segreto *et al.* (2013)).

Signal processing is the next step after choosing appropriate sensors and signal acquisition to magnify the effect of monitoring parameters by removing noises. Time domain analysis, frequency domain analysis and time-frequency domain analysis are three common approaches for this step (Rehorn *et al.* (2005)). While many researches are devoted in time domain and frequency domain analysis due to low complexity, time-frequency analysis is well-suited for this application as it examines both time variant and frequency characteristics of the signal simultaneously. Based on the non-stationary nature of faulty signals, time-frequency analysis can provide discriminative information about machinery health conditions. Therefore, discriminative fault features can be extracted from a faulty signal by choosing a proper time-frequency method (Feng *et al.* (2013)). Rehorn *et al.* utilized s-transform as a time frequency transformation method, and proposed a time-frequency domain feature, selective regional correlation, for machining condition monitoring (Rehorn *et al.* (2006)). In another research, a comparative study is perfomed between five time-frequency transformation methods for the purposes of TCM in milling operation (Rad *et al.* (2014a)).

The signal representation in time-frequency domain has high dimensions. Therefore, after time-frequency step, dimensionality reduction methods are useful. Dimensionality reduction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) are popular among the literature to perform this task (Elgargni *et al.* (2015); Shi & Gindy (2007); Jin *et al.* (2014)). Spectral subtraction is another method which can be implemented to enhance signal quality. It it originally used in speech enhancement to remove the effects of steady sounds in the environment (Bodin & Villemoes (1997); Boll (1979)). Similar to the sound and speech analysis applications, this method is employed as a noise reduction tool. Fault diagnosis applications can use it to reduce the steady state part of the signal and present fault characteristics. For example, El Bouchikhi *et al.* proposed an algorithm for fault diagnosis of induction machine bearings using spectral subtraction method. In this study, stator current frequency response of the healthy machine is subtracted from spectrum of machine's current acquired signal to present better fault indicators (Choqueuse *et al.* (2013)).

During the machining process, many parameters such as operational conditions, depth of cut, feed rate and workpiece material are changing which may degrade the monitoring performance and can reduce system robustness. Moreover, the relation between signals and monitoring parameters are often non-linear and complex. Therefore, powerful methods are required to perform the decision making task. Many methods such as artificial neural networks (ANN), support vector machine (SVM) and Bayesian networks are employed to perform this task in the literature. Patra *et al.* investigated tool wear during micro drill using thrust force signals and ANN method (Patra *et al.* (2017)). In another study, discrete wavelet transform (DWT) and SVM are used along with sound signals for tool condition monitoring in face milling (Madhusudana *et al.* (2017)). Tobon-Mejia used Baysian network method for TCM and the estimation of its remaining useful life (RUL) in machining process (Tobon-Mejia *et al.* (2012)).

Recently, powerful characteristics of deep learning methods draw attention of researchers in different fields and helped them to solve many challenges in machine learning domain (Schmidhuber (2015)). Deep learning refers to machine learning techniques with deep architectures and multiple layers which enable them to learn highly complex relationships from even low-processed to raw signals (Deng (2014)). In an era in which sensors are actively producing high amounts of data, such techniques are able to make the most information out of the big data. They are therefore less dependent on specific applications and frameworks and have powerful characteristics to outperform other methods when the relationship between the input data and desired output are complex (Jia *et al.* (2016)). Despite their high potential, they are relatively recent in the field of machinery condition monitoring. Zhao *et al.* employed Long Short-Term Memory networks (CBLSTM) for tool condition monitoring in milling process (Zhao *et al.* (2017)). Jing *et al.* developed a CNN based method for gearbox condition monitoring using frequency data of vibration signals and their method outperformed some of the common machine learning algorithms (Jing *et al.* (2017)). In another study, vibration signals of a gerbox system are preprocessed using statistical measures from the time domain and frequency band energy from frequency domain. Then the feature vector is fed to CNN to train it to detect gearbox faults (Chen *et al.* (2015)). Based on high potentials of deep learning methods, further research is crucial to examine them with different signals and levels of signal processing in TCM applications.

In this study, a TCM system is proposed using convolutional neural network as a powerful and established deep learning method. In the first step, force signals and vibration signals from ETS experimental dataset are selected independently to develop the monitoring system. Wavelet packet transform is employed for signal processing step to transform the signal to time-frequency domain. The final step is the machine learning algorithm which the proposed method is compared with three common methods in the literature, support vector regression, Bayesian rigid regression and Nearest neighbor regression methods. In the next step of the study, spindle current signal from Nasa_Ames dataset Agogino & Goebel (2007) is used for further validation of the method. Spectral subtraction is an ideal candidate to process current signals as it is power based and not a vectorial format in contrast to force and vibration. . Spectral subtraction method is applied around tooth path frequency. Afterwards, further noise processing is performed on the output of spectral subtraction and a number of features are generated to represent the fault in the signals. Finally, the comparative study between the machine learning algorithms is performed to see the results with a different dataset, signal and higher levels of signal processing.

This paper is organized as follows: Section 2 represents the backgrounds and formulation of the algorithms which are used in this paper. The proposed algorithm is explained in Section 3. Section 4 introduces the two datasets which are used in this study for validation of the work. Results and discussion are presented in Section 5 and Section 6 is dedicated to conclusion.

4.2 Background of methods

This section presents the formulation and background of the algorithms and techniques which are used in this paper.

4.2.1 Wavelet Transform

Wavelet transform is one of the methods that is widely used for fault diagnosis and health condition monitoring. The main difference between wavelet transform(WT) and Fast Fourier Transform (FFT) is that in wavelet transform, wavelets are used as the basis instead of sinusoidal functions that are used in fast Fourier transforms. It is an effective tool for transient signal analysis as well as time-frequency localization since it adds a scale variable in addition to the time variable in the inner product transform. It has a better time localization but a lower frequency resolution for higher frequency components. In contrast, for lower frequency components, the frequency resolution is higher while the time localization is worse. Following equation describes the formulation of the continuous wavelet transform (Feng *et al.* (2013)).

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\, \psi \frac{(u-t)}{a}\, du \tag{4.1}$$

where wavelet $\psi(u-t)/a$ is derived by dilating and translating the wavelet basis $\psi(t)$, and $1/\sqrt{a}$ is a normalization factor to maintain energy conservation and $a > 0$.

4.2.2 Spectral Subtraction

Spectral subtraction is a method which was originally used for speech signal enhancement. A signal is considered a combination of noise and clean speech, therefore the noise spectrum is estimated during speech pauses, and an estimation of the noise spectrum is subtracted from the noisy speech spectrum to obtain clean speech. It can be used in fault diagnosis applications by removing the steady state and normal process part of the spectrum from new signals to obtain their anomalies and fault signatures. Consider a measured signal which consists of the steady state normal component and additive fault (Boll (1979); Choqueuse *et al.* (2013)):

$$y[n] = s[n] + d[n] \tag{4.2}$$

where y[n],s[n] and d[n] are the sampled measured signals, fault and steady state component, respectively. The frequency domain representation of the signal is given by:

$$Y(jw) = S(jw) + D(jw) \tag{4.3}$$

Therefore, the fault component of the signal can be obtained based on the following equation:

$$\widehat{S}(jw) = Y(jw) - \widehat{D}(jw) \tag{4.4}$$

where $\widehat{S}(jw)$ is the fault related spectrum estimate and $\widehat{D}(jw)$ is an estimate of the steady state component of spectrum. $\widehat{D}(jw)$ is often obtained using the time-averaged signal spectrum using the normal healthy state of the system:

$$\widehat{D}(jw) \cong |\overline{D}(jw)| = \frac{1}{K}\sum_{i=0}^{K-1} |D_i(jw)| \tag{4.5}$$

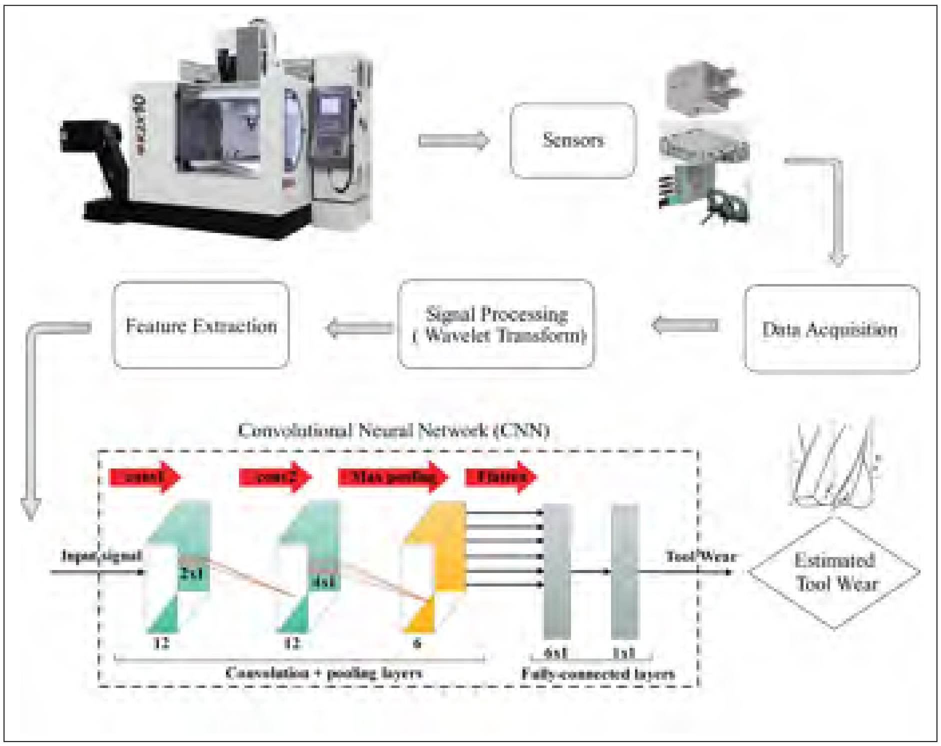

Figure 4.1 The monitoring system framework

4.2.3 Convolutional Neural Network

Deep convolutional neural networks (CNNs) have recently demonstrated a great success in many machine learning tasks, such as regression, prediction, etc. Such models have been exploited to appropriately characterize internal variations (intra-class) within a large amount of data. The special characteristic of this network is that the network learns data-driven filters to convert the data to features that describe the inputs and represent variables of interest inside the network which are usually performed separately in traditional methods (Bouvrie (2006)). Therefore, it can achieve high performance even with minimal preprocessing. The general architecture of CNN consists of one input layer, one or multiple convolutional layers, pooling layers, fully connected layers and one output layer.

In a convolution layer , the feature maps from previous layers are convolved with learnable kernels and fed to the the activation function to construct the output feature map. Each output map may combine convolutions with multiple input maps. The general formulation for a convolutional layer is (Bouvrie (2006)):

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{4.6}$$

where f is a nonlinear activation function, M_j represents a selection of input maps, l is the index for each convolution layer, K is a square matrix with the size of kernels and b is an additive bias given to each output map.

Pooling layers are generally used after the convolutional layers to produce down sampled versions of the input maps. Therefore, the number of output maps will be the same as the number of input maps, but their dimensions are decreased. In terms of formulation (Bouvrie (2006)):

$$X_j^l = f\left(\beta_j^l down(X_j^{l-1}) + b_j^l\right) \tag{4.7}$$

where $down()$ represents a sub-sampling function. Max pooling is an example of such functions which uses the maximum value from each cluster of neurons at the prior layer (Ciregan *et al.* (2012)). Each output map has its own multiplicative bias β and an additive bias b.

Finally, fully connected layers which are traditional multi-layer perceptions (MLPs) (Ruck *et al.* (1990)) compute the desired outputs from the neurons of the previous layers.

4.3 Proposed Methodology

In this section proposed methodology of this paper is elaborated. The first step of the system is signal acquisition. Three different sensors (dynamometer, accelerometer and current sensor) are examined using two different datasets. The framework of the system with its different steps is depicted in Figure 4.1. After the data acquisition step, signals are processed to extract fault indicators and remove noise.

Time-frequency transformation is used for this task due to its promising capability in revealing the time variant characteristics of the signals in frequency domain using Morlet wavelet transform method. We favored the Morlet wavelet in this research as in mechanical dynamical signals, impulses are usually the symptoms of faults and the Morlet wavelet is very similar to impulse component Lin & Qu (2000). Furthermore, Jauregui *et al.* in their research on frequency and time-frequency analysis of cutting force and vibration signals for tool condition monitoring, reported the Morlet wavelet function as a good candidate for the feature extraction applications, as it provides a good balance between time and frequency resolutions Jáuregui *et al.* (2018).

The next step is to extract features from the wavelet transform that describe the fault properly. Wavelet packet transform is employed as it permits decomposing signals into uniform frequency bands Barros & Diego (2006). WaveletPacket function from PyWavelets Wasilewski (2010), the scientific Python module for Wavelet Transform calculations is used to perform the transformation. The algorithm proposed in this paper uses the morlet as the wavelet function and with 4 as the levels of decomposition. The output of the transformation is divided into 16 uniform bands. The rms value of each grouped output frequency band is obtained and the first 12 bands rms values is used as the input to the machine learning without further processing. Therefore, minimum pre-processing is implemented to explore the capability of CNN.

Contrary to other hand-crafted feature learning models, these data-driven models are capable of learning discriminative non-linear feature representations. Thus, they can provide an effective prediction tool for fault detection by learning robust feature representations directly from the input signals.

A deep CNN model is proposed in this paper to accurately predict the faults in machining process. To that end, a simple yet effective architecture as shown in Figure 4.1 is considered due to the constraints of tool condition monitoring system. The proposed architecture is comprised of two convolutional layers (conv1 and conv2), where each layer has 12 kernels of 2x1 and 4x1, respectively. To keep the original signal size and avoid decreasing the number of features, the stride is considered as 1 and these convolutional layers are designed consecutively. A pooling layer is used after those successive convolutional layers to handle the spatial size of the feature representation through max pooling, as well as, to control the number of parameters (computational complexity of the network) and overfitting. The output of pooling layer is first flatten and then fed into two fully-connected layers. The fully-connected layers are responsible to compute the softmax activation with a matrix multiplication followed by a bias in order to produce the prediction value.

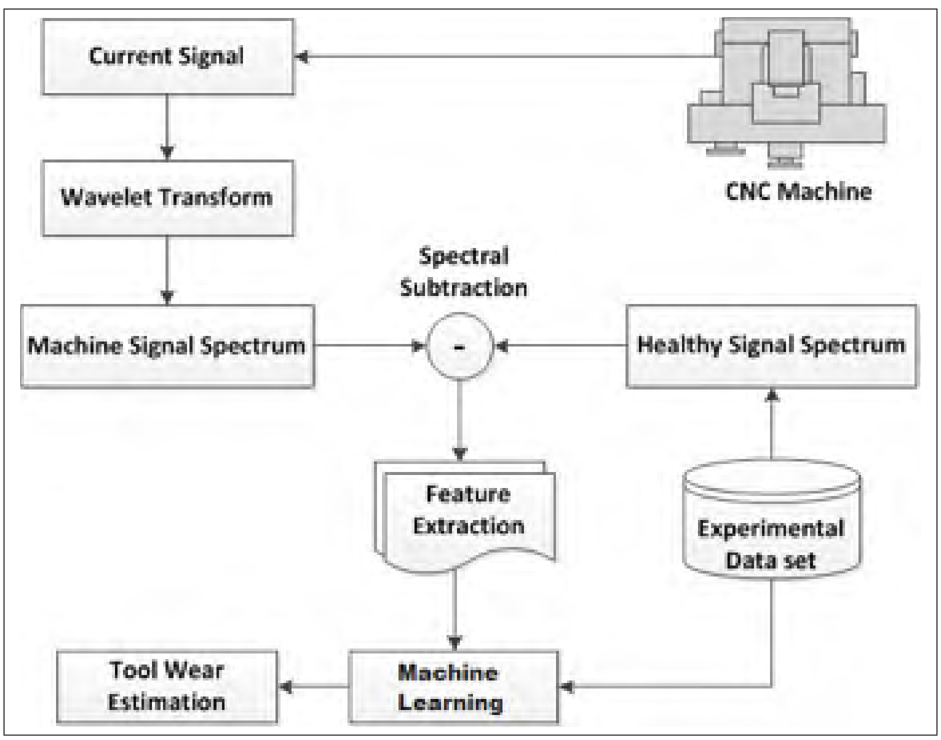

Figure 4.2 Spectral subtraction method for current signal

For the system which uses current signals, spectral subtraction method is applied to the signals. For this purpose, a local average of the spectral magnitude in different frequency bands is extracted using the dataset. For each new signal, the estimated healthy spectrum under the same cutting conditions is subtracted from the signal. Figure 4.2 illustrates the diagram of spectral subtraction method. Further noise canceling and signal refinement is performed after the spectral subtraction step.

In the next step, various features are extracted from the output of spectral subtraction step. Maximum energy and frequency of occurrence, variance, standard deviation and width of frequency response are among the extracted features. Afterward, CNN model which is used for this step is similar to the one which was explained in Figure 4.1.

4.4 Experimental Datasets

4.4.1 ETS Dataset

A set of experiments are performed to measure tool flank wear during machining of hard to cut materials. K2X10 Huron high speed CNC machine of the LIPPS laboratory at ETS is used to perform the experimental tests. A multi-component Kistler dynamometer 9255-B, coupled with charge amplifiers, was used to measure the cutting forces in three orthogonal directions (Fx/Fy/Fz). A tri-axial accelerometer was mounted on the spindle of the machine with a sensitivity of 100mV/g for measuring acceleration.

D2 high speed tool steel is selected as the workpiece material with hardness of 60-62 HRC due to its high wear resistance in order to investigate tool wear in machining hard material with dimension of 200 mm $\times$ 54 mm $\times$ 4 mm. Carbide Walter End Mill Protostar H50 Ultra tool with 6 teeth is selected as the cutting tool with 50 degrees of helix angle. Different cutting speeds of 2500 rpm and 6000 rpm and feed rates of 0.12 mm/tooth and 0.05 mm/tooth with 4 mm depth of cut and tool wear were measured at different intervals which results in 63 cases with different tool wears and cutting conditions. Figure 4.3 demonstrates this experimental setup.

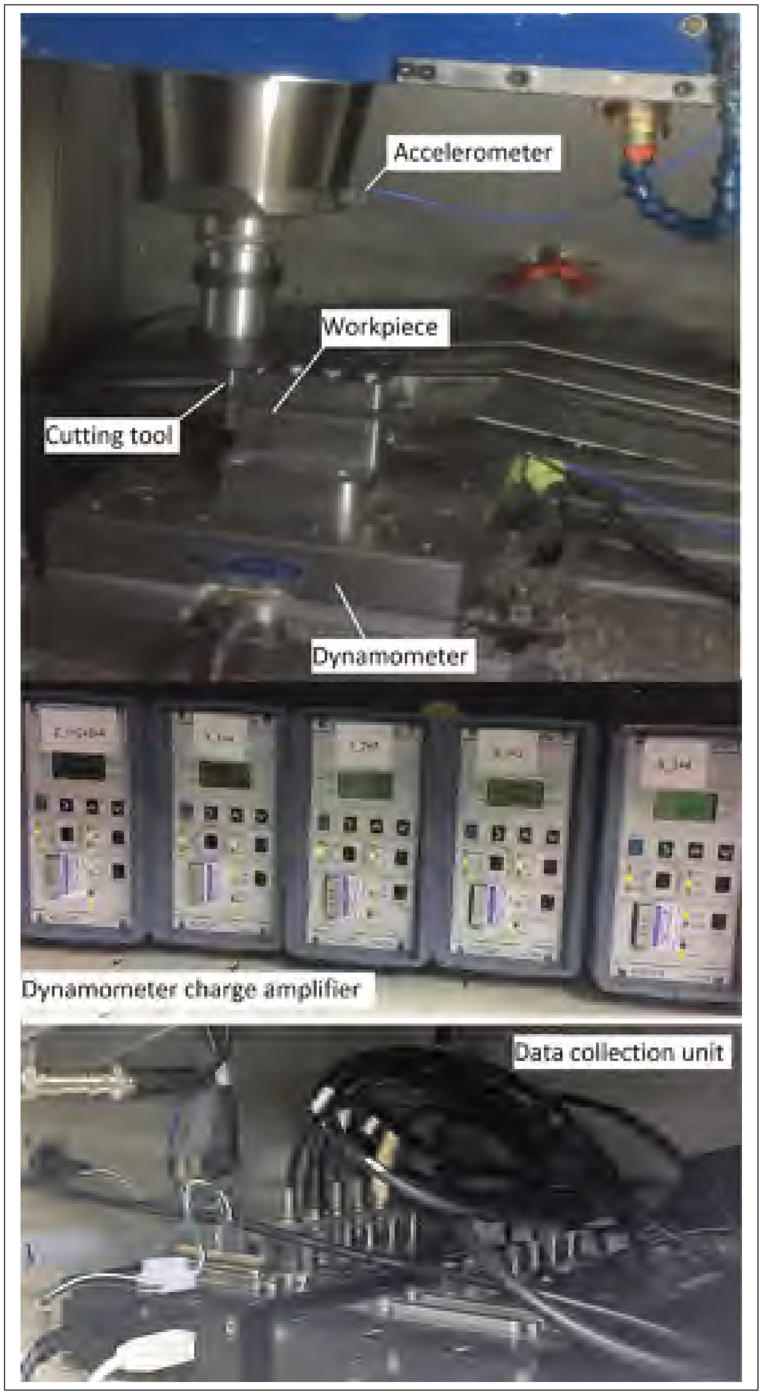

Figure 4.3 Experimental set up

4.4.2 Nasa_Ames Dataset

Tool fault detection models and validation of the research method is implemented by the benchmark NASA Ames and UC Berkeley milling dataset Agogino & Goebel (2007). The experiments are performed under various operating conditions using the Matsuura MC-510V machining center. In this research, spindle current sensor is selected based on its ease of use and practicality in industrial application. The dataset signals include cases with changes in depth of cut, feed rate and therefore, the effect of these parameters can be investigated on the monitoring system accuracy and system will be developed under varying cutting parameters.

The tool is a 70mm face mill with 6 KC710 inserts based on its industrial applicability. Workpiece material in the research is cast iron. An OMRON K3TB-A1015 current converter feeds the signal from one spindle motor current phase into the cable connector and a model CTA 213 current sensor(Flexcore Div. of Marlan & Associates, Inc.) is used for data acquisition. Flank wear (VB in μm), which is defined as the distance from the cutting edge to the end of the abrasive wear on the flank face is considered as the fault and its value is reported in all the experiments using a microscope.

4.5 Results and discussion

In this section, results for the three monitoring systems are presented. The first one uses force signals from the ETS dataset as the monitoring signal. The second one employs the vibration signals from ETS dataset and finally the third subsection investigates the system with current signals from the Nasa_Ames dataset and spectral subtraction method.

4.5.1 Tool wear estimation using force signals from ETS dataset

The methodology of this system is depicted in Figure 4.1. This system uses force signal as the monitoring indicator from the ETS dataset. The data is divided into two categories, training and testing. The system is trained using the training subset which consists of 70% of the data. Afterwards, the system is tested with 30% of the data. The feautures and tool wear are normalized before machine learning step and denormalized after tool wear prediction. MinMaxScaler function from Scikit-learn python library Pedregosa *et al.* (2011) is employed as an standard machine learning feature normalization method to normalize both inputs and tool wear. It trains an estimator using a linear scaling function to transform features .This estimator scales and translates each feature individually such that it is in the given range (between zero and one in this paper) on the training set using a linear interpolation.

Table 4.1 Comparison between different machine learning algorithms (force)

Regression Algorithms	Average Accuracy %	Root Mean Square Error
Bayesian Ridge Regression	73.1	0.1815
Nearest Neighbors Regression (KNN)	71.5	0.2021
Support Vector Regression (SVR)	79.0	0.1103
Convolutional Neural Network (CNN)	88.2	0.0709

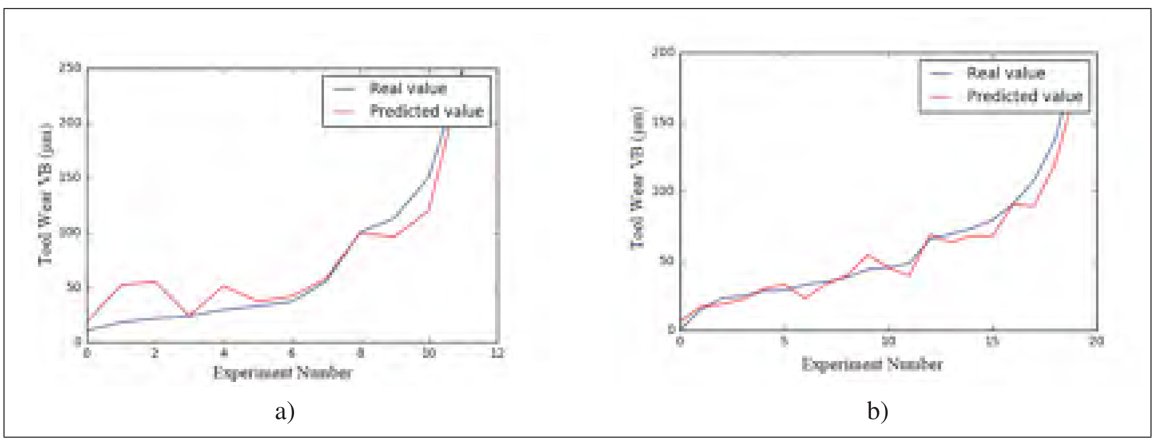

Figure 4.4 Estimated and real tool wear values using force signals a) First cutting, b) Second cutting tool

The keras deep learning library is employed Chollet (2015) with tensorflow as the back-end Abadi *et al.* (2016) to implement the proposed CNN model. Three of the common machine learning methods (SVR, KNN and Baysian network) in this field are also implemented using Scikit-learn machine learning library Pedregosa *et al.* (2011) as a baseline to compare the performance of the CNN based system with these methods. For the monitoring system, average accuracy in percentage (the differences between predicted and actual tool wear value divided by average of tool wears) and RMSE are calculated as representative of the performance from the Scikit-learn machine learning performance analysis toolboxes.

Table 4.1 presents the results of tool wear estimation using test dataset for different machine learning algorithms. Based on the results, CNN has the highest accuracy (88.2%) and lowest root mean square error(RMSE) (0.0709) which are acceptable for most industrial applications. It can be observed from the results is a relatively high difference between the accuracy of CNN and other methods due to the fact that signals were not processed after the time-frequency transformation method. Convolution layers of the CNN was able to filter the data and convert it to more discriminative features. Figure 4.4 presents the predicted versus actual tool wears using the CNN based algorithm for two tools from the no wear state up to the high tool wear values. For each tool, the experiments start with no wear (VB=0) and the curves shows gradual increase in the tool wear as cutting is continued until high tool wear values. Based on the figure, the estimated tool wear greatly correlates with actual tool wear.

Table 4.2 Comparison between different machine learning algorithms (vibration)

Regression Algorithms	Average Accuracy %	Root Mean Square Error
Bayesian Ridge Regression	66.9	0.2251
Nearest Neighbors Regression (KNN)	57.0	0.281
Support Vector Regression (SVR)	59.6	0.2701
Convolutional Neural Network (CNN)	84.6	0.086

4.5.2 Tool wear estimation using vibration signals from ETS dataset

The system in this section is similar to the force based monitoring system of the previous section, except that the force signal is replaced with the spindle vibrations. The ETS dataset is used for this system as well. Data is divided into two categories, Training and testing with 70% and 30% of the data respectively. The keras deep learning library is employed to implement the proposed CNN model and Scikit-learn for othe machine learning methods similar to previous sections.

The result of tool wear estimation using vibration test dataset is provided in table 4.2. The accuracy of CNN is 84% which is slightly lower than the force-based system of previous sections, but still promising for real-world applications. For the vibration-based systems, there is a high difference between the accuracy of CNN and other machine learning methods. It can be interpreted that it is due to the fact that vibration signals do not provide features directly related to tool wear without specific and hand-crafted pre-processing. However, convolution filters in CNN architecture were able to convert the data to discriminative features. RMSE value of CNN (0.086) is also significantly lower than other methods. Figure 4.5 illustrates the predicted versus actual tool wear using the CNN based algorithm for two different tools. Experiments start with a new tool with no wear (Experiment 0) state up to the high tool wears.

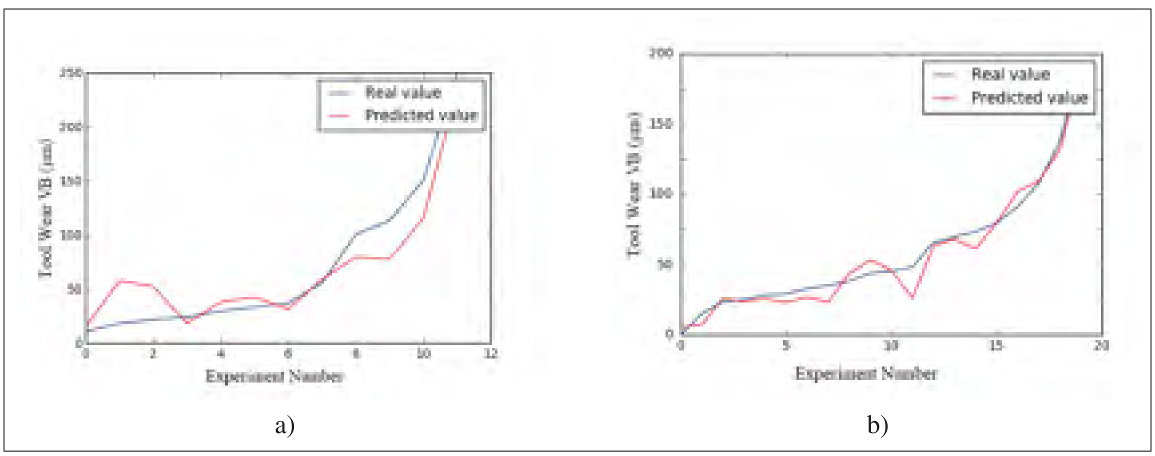

Figure 4.5 Estimated and real tool wear values using vibration signals, a) First cutting tool, b) Second cutting tool

4.5.3 Tool wear estimation using spindle current signals from Nasa_Ames dataset

Table 4.3 Comparison between different machine learning algorithms (current)

Regression Algorithms	With Spectral Subtraction		Without Spectral Subtraction	
	Average Accuracy %	RMSE	Average Accuracy %	RMSE
Bayesian Ridge Regression	85.8	0.091	79.6	0.188
Nearest Neighbors Regression (KNN)	79.1	0.220	78.5	0.232
Support Vector Regression (SVR)	85.5	0.102	80.1	0.181
Convolutional Neural Netwoek (CNN)	87.2	0.088	81.5	0.156

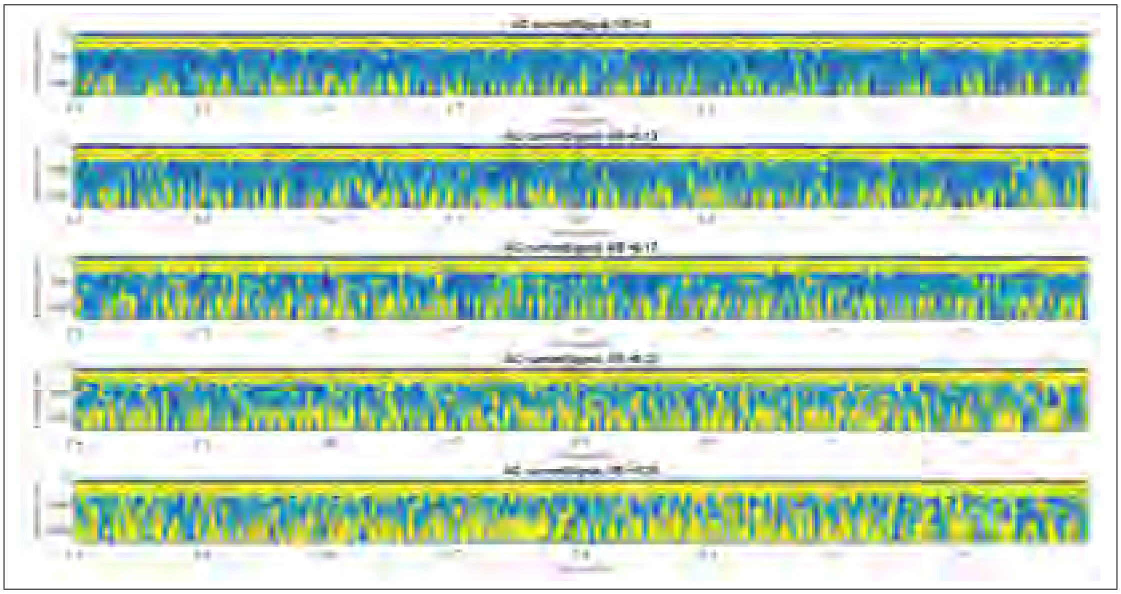

Figure 4.6   Spindle current signals with different tool wears in wavelet domain

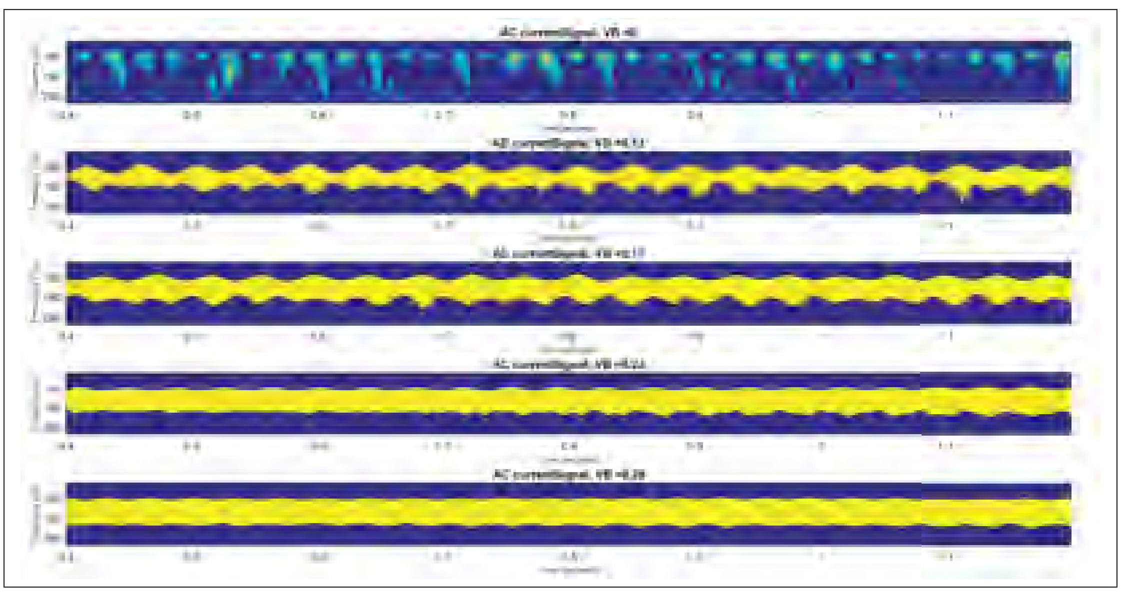

Figure 4.7   Spindle current signals with different tool wears after spectral subtraction

The last section of the study investigates a system design based on the spindle current signal as the fault indicator. Nasa_Ames dataset is utilized for the validation. Two systems are trained for comparison one with spectral subtraction method and another without this method. The architecture of the first system is illustrated in Figure 4.2. The architecture of the second method is similar to the force and vibration based systems, which directly fed machine learning with wavelet transform based features. Therefore, the effect of higher pre-processing is investigated in the performance of the algorithms.

Figure 4.6 depicts the result of wavelet transform using some sample signals from the dataset. The diagram represents the WT output for the healthy signal (VB=0) as well as four states of the fault. It can be observed from the diagrams that the signal has higher energy around tooth pass frequency. As the fault value increases, the magnitude and density of wavelet are increased. However the data is still noisy and needs to be further processed to extract discriminative features. The processed data after the spectral subtraction is shown in Figure 4.7. For this analysis, an average estimation of spectrum of the current signal around tooth path frequency for healthy case is extracted from the dataset. For each new signal of the machine, the estimated healthy spectrum under the same cutting conditions is subtracted from the signal. It normalizes the signals based on their cutting conditions and magnifies the effect of tool wear. Comparing the Figure 4.6 and Figure 4.7, the signal quality is enhanced after the spectral subtraction and the signals are more discriminative with respect to tool wear.

The CNN architecture and machine learning implementation for this step is similar to two previous systems. Figure 4.8 reports the loss values converging close to zero during the epochs of the training step. Afterwards, the unseen test dataset signals are fed to the system.

Based on the results of Table 4.3, CNN is superior to other methods with 87.2% accuracy in tool wear estimation for the system with spectral subtraction Bayesian rigid regression and support vector regression methods also have satisfactory results of 85.8% and 85.5% respectively. Based on the comparison between the systems with and without spectral subtraction, spectral subtraction increased the accuracy of each algorithm by approximately 5%. Therefore, it can be concluded that although CNN has powerful capabilities to interpret data with minimum preprocessing, it would still benefit from advanced and efficient signal processing methods, specially under limited number of training samples.

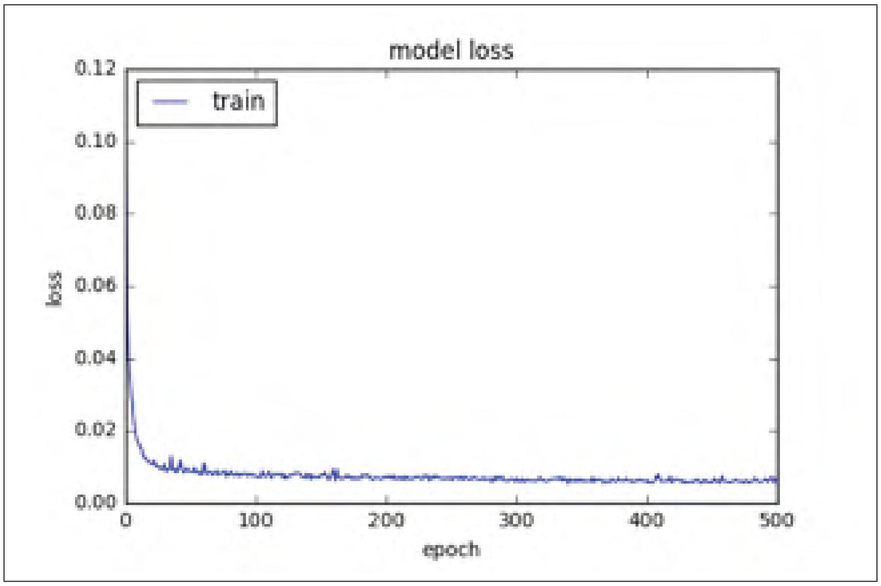

Figure 4.8 Loss function during training process

Based on the results of three studied systems, CNN consistently provided higher results than other algorithms with different datasets and signals which proves its efficiency. The major advantage of CNN is in lower processed signals which has ability to extract relevant information compared to other methods.

4.6 Conclusion

In this research, a tool condition monitoring methodology is proposed and tested under changing cutting parameters. Force and vibration signals from the ETS dataset and spindle motor current signals from Nasa_Ames dataset are used as monitoring signals. Wavelet transform as an advanced time-frequency transformation method is used in the signal processing step due to its great applicability to process signals and reveal rich information in both time and frequency domain simultaneously. A deep CNN method is also implemented as the last step to model the complex relationships between extracted features and tool wear values.

Wavelet analysis revealed the time variant characteristics of frequency response of the signal and the study confirms its performance and applicability for tool wear monitoring. Spectral subtraction method is employed for the current signal which significantly improved its condition and magnified the signature of tool wear by removing the steady state part of the signal due to normal cutting and magnified the remaining fault characteristics.The comparison between the systems with and without spectral subtraction shows approximately 5% increase in the accuracy of systems which benefit from this algorithm.

Tables 4.1-4.3 report the comparative results of the CNN-proposed methodology of the paper with some of the common machine learning techniques. Based on the results, CNN consistently outperforms other machine learning algorithms between all three signals from two datasets which proves its robustness and high performance in this application. CNN improved the accuracy of force and vibration based methods significantly (around 15%). It is interpreted as the results of CNN convolution layers which filtered the signal and extracted discriminative features from the raw wavelet response. Therefore, it is beneficial in reducing the cost of specific engineering data manipulations and improving accuracy in the case of low quality signals.

Acknowledgement

The authors gratefully acknowledge the experimental mill data provided by UC Berkely BEST lab and NASA Ames Prognostic Data Repository.

CHAPTER 5

TOOL CONDITION MONITORING METHOD IN MILLING PROCESS USING DEEP TRANSFER LEARNING

Fatemeh Aghazadeh[1], Antoine Tahan[1], Marc Thomas[1]

[1] Département de Génie Mécanique, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Under Review in INT J ADV MANUF TECH, January 2020

Abstract

Online condition monitoring is an important step toward achieving total autonomy for manufacturing plants. It reduces routine checks, enables proactive maintenance and improves productivity. The widespread availability of low-cost sensors has led to a significant increase in the popularity of data-driven machine learning techniques in condition monitoring applications. In particular, deep learning algorithms have recently been receiving a lot of attention within the diagnosis and prognosis communities due to their exceptional performance in exploiting information to solve complex non-linear problems. Notwithstanding the advantages of machine learning algorithms, one of their main drawbacks is their heavy data requirements. Furthermore, knowledge is not easily transferable between related systems. For example, a monitoring system trained using data from a specific machine and task is not reusable on another machine or a different task. Therefore, for each machine and task, considerable training data is required. Transfer learning is an advanced approach, which can tackle this issue. It refers to reuse of a trained machine learning model on a given problem for a related yet new problem. The present research focuses on leveraging transfer learning in tool wear monitoring application to increase the scalability of machine learning in condition monitoring while benefiting the accuracy of these methods. To validate this approach, two different datasets, with different machining centers, cutting tools and workpiece materials are used. The NASA-Ames tool wear benchmark dataset is used in the first step to train a deep learning based tool wear estimation model. The benchmark dataset comprises two different workpiece materials and

numerous cases of tool wear under varying cutting parameters. Spindle vibration signal is used to indirectly indicate tool wear characteristics. Wavelet Transform is employed for signal processing to simultaneously reveal both the time domain and frequency domain features of the signal. Deep convolutional neural network (CNN) method is leveraged to model the complex relations between tool wear and vibration signal. After training a model using the first dataset to detect the tool wear, this model is used as the source model for the second dataset. The second dataset is experimentally acquired with a K2X10 Huron high speed CNC machine in the TS LIPPS and Dynamo labs. The convolutional layers of the first (source) model fitted on the source task are used as pre-trained model for the second model, with a lower number of experiments. Therefore, only fully connected layers are retrained to fine-tune the second model to adopt to the new domain. The results prove that with the proposed technique, robust monitoring systems can be achieved with much lower data requirements by benefiting from the knowledge of a pre-trained model.

5.1 Introduction

Machining processes are among the main components of today's industrial manufacturing environment, which demands higher productivity, production quality, worker safety and lower operational costs. Machining operation automation is a key factor in achieving these requirements. In the context of machining automation, online process monitoring is crucial in ensuring production safety and quality. Tool wear is one of the most common and costly defects of the machining process. It is caused by excessive contact forces and friction between the cutting tool and workpiece material, high temperatures at the cutting surfaces and pressure of chips on the tool. Tool wear can deteriorate the surface finish or cause damage to the tool, workpiece or to the machining center if is not detected and repaired on time (Zhu & Vogel-Heuser (2014)). Developing a robust tool condition monitoring (TCM) system is thus invaluable when it comes to increasing the productivity and quality of the machining process. A practical and relatively cost-effective approach for designing a TCM system involves using physical parameters such as vibration, acoustic emission, force and power signals to monitor indirectly the state of the system (Snr (2000)). Such techniques, which are known as indirect methods, can be used to fulfill TCM requirements; they represent an alternative to directly measuring tool wear, otherwise known as the direct method. The direct approach may cause disruptions to the machining process, and is often more expensive to implement. Another advantage of indirect methods is that the same sensor used to identify tool wear can also be used for other monitoring tasks (Abellan-Nebot & Subirón (2010)).

Force is an indicator that can reveal fault signatures with high accuracy. In a study, Zhu *et al*. (Zhu & Zhang (2019)) proposed a generic wear model using force signals. The study investigates the relationship between milling force and tool wear to establish a technical foundation for on-line force-based wear monitoring. The force signal is highly sensitive to tool wear, which thus makes it desirable for wear monitoring. However, it is also highly dependent on other operating conditions and costly for practical applications (Abellan-Nebot & Subirón (2010)). Vibration signals, on the other hand, provide an acceptable combination of practicality and accuracy in this application. Aghdam *et al*. (Aghdam *et al.* (2015)) studied the correlation of the tool holder assembly vibration and tool flank wear in turning operations. In their study, a set of features are extracted from the recorded signals using autoregressive moving average (ARMA) model and are correlated to tool wear. Acoustic emission signal is another practical fault indicator which is widely used in the literature (Rad *et al.* (2014b)). Power (and current) signals are low cost fault indicators that can produce accurate monitoring results, particularly when combined with other signals. In a study by Soltani Rad *et al*. (Rad *et al.* (2013)), spindle current signals from a milling machine is used to investigate the tool wear. S-transform is employed to transform the signals into the time-frequency domain in the signal processing step. Combining multiple sensors in a complimentary way (Sensor fusion) is another approach for tool wear monitoring. It is employed to increase accuracy and reliability of the monitoring system. In a study, Wang *et al*. combined information of the vibration and force sensors using a multi-scale principal component analysis (MSPCA) method for tool condition monitoring in milling operation (Wang *et al.* (2019)).

Signal processing is the next task to remove noises and emphasize the signature of tool wear (Rehorn *et al.* (2005)). Time-frequency transformation is a powerful signal processing tool for this application as it investigates both time and frequency domain characteristics of the signal simultaneously (Feng *et al.* (2013)). For example, Wavelet transform is employed by Benkedjouh *et al*. (ref. date) in combination with the blind source separation (BSS) technique to predict the remaining useful tool life in milling operation (Benkedjouh *et al.* (2018)). Elsewhere, Soltani Rad *et al*. conducted a comparative study among common time-frequency transformation methods for tool condition monitoring in the milling process (Rad *et al.* (2014a)).

Machine learning algorithms are attractive among researchers for modeling the relations between monitoring signals and tool wear, especially under varying cutting conditions, such as feed rate and depth of cut. This is mostly due to superior capabilities of machine learning when it comes to finding non-linear patterns within datasets. These time-frequency transformation methods include the artificial neural network (ANN), hidden Markov model (HMM), support vector machine (SVM), fuzzy logic and other regression methods (Zhou & Xue (2018)). Deep learning algorithms have recently been receiving a lot of attention in condition monitoring, as well as in many other fields, because of their exceptional capacity to learn complex patterns (Schmidhuber (2015)). Deep learning is a subset of machine learning, with a deep architecture of multiple layers, which enables it to learn highly complex models even from low-processed to raw signals (Deng (2014)). In a study by Aghazadeh *et al*. (Aghazadeh *et al.* (2018a)), force and vibration signals are independently subjected to time-frequency transformation and fed to convolution neural networks (CNNs) to estimate the tool wear in the milling process. Moreover, the spectral subtraction method is applied to current signals and the output is fed to CNNs for tool wear prediction . Luo *et al*. proposed a method for early fault detection leveraging a deep learning model consisting of SAE and BPNN layers to automatically select the impulse responses from the vibration signals. Afterwards, dynamic properties are identified from the selected impulse responses to help detect mechanical faults (Luo *et al.* (2019)). In another study, Zhao *et al*. employed Convolutional Bi-directional Long Short-Term Memory networks

(CBLSTM), which uses CNN to extract local features from the raw signals and bi-directional LSTM to encode temporal information for tool wear prediction (Zhao *et al.* (2017)).

Despite the high accuracy of machine learning and deep learning algorithms in solving complex problems, they have certain disadvantages. Their relatively expensive data requirements especially in condition monitoring applications where experimental data acquisition can be costly makes their usage challenging and unscalable. Moreover, most of them only work well under a common assumption, namely, that the training and test data should come from the same feature space, with the same distribution (Pan & Yang (2010)).

This means a model trained on a specific machine and for a certain task, is not reusable for another machine or different tasks and conditions. Transfer learning is emerged in recent years as a new learning framework to address these drawbacks. Transfer learning method refers to the act of gaining knowledge while solving a problem and applying it to solve a different related problem (Pan & Yang (2010)). It is widely exploited in computer vision applications by fine-tuning pre-trained deep learning models from large image datasets for use in other image recognition tasks. Shin *et al.* achieved superior results by fine-tuning an ImageNet-based pre-trained model in medical imaging specifically for cases with insufficient amounts of labeled training data (Shin *et al.* (2016)). It is also commonly used in natural language processing to gain knowledge from large available datasets, which is then used in another model. Baziotis *et al.* proposed a Bidirectional LSTM method for multiple tasks, such as emotion classification and intensity regression on tweets-related data. They employed a set of word2vec word embeddings trained on a large collection of 550 million Twitter messages. Subsequently, they pre-trained the Bi-LSTMs on the dataset of the Semeval 2017 competition and fine-tuned it for some new tasks(Baziotis *et al.* (2018)). Transfer learning can add tremendous value to the condition monitoring field, given the difficulties inherent in preparing large experimental labeled datasets. Guo *et al.* proposed a deep convolutional transfer learning architecture consisting of two main steps: condition recognition and domain adaptation. The first is used to automatically learn features and recognize the health conditions of machines, while the latter facilitates the first step in learning domain-invariant features by maximizing domain recognition errors and minimizing the probability distribution distance. They validated the transfer learning approach with multiple datasets (Guo *et al.* (2018)).

Based on the literature review and transfer learning results in other similar fields, there is a great potential to apply this approach to tool condition monitoring to achieve acceptable accuracy with considerably lower data requirements.

In the present study, a tool condition monitoring system is proposed using a deep convolutional transfer learning method. It applies a Wavelet transform algorithm on vibration signals from the milling process to reveal both time and frequency domain characteristics of the signals. Afterwards, a set of features is obtained using the frequency band energies calculated in the previous step. Finally, a deep convolutional neural network method is trained on the extracted features as inputs to estimate the tool wear. This research uses two datasets for validation. The NASA-Ames benchmark milling dataset is used to pre-train a model with a large subset of data from different workpiece materials and operation conditions, after which a second dataset containing experimental data from the LIPPS and Dynamo ÉTS labs is used to design the target model. The second system leverages the weights and information of certain layers from the pre-trained model and retrains certain elements to adopt the model to new domain . This paper is organized as follows: Section 2 presents the formulation and backgrounds of the techniques employed in the paper. The proposed methodology is explained in detail in section 3. Experimental datasets are presented in section 4. Section 5 represents results and discussions, and section 6 concludes the paper.

5.2 Background of Methods

5.2.1 Wavelet Transform

Wavelet transform is a signal processing algorithm that is widely used for fault diagnosis and health condition monitoring. Wavelet transform decomposes a signal into different ranges of frequencies using wavelets as the basis; these wavelets act as low-pass and high-pass filters.

The main difference between Wavelet transform (WT) and Fast Fourier Transform (FFT) is that Wavelet transform uses wavelets instead of sinusoidal functions, as is the case in fast Fourier transforms. Wavelet transform has a better time-frequency localization as it introduces a scale variable alongside a time variable in the inner product transform. For higher frequency components, WT has a better time localization but a lower frequency resolution, while with lower frequency components, the frequency resolution is higher and the time localization is decreased. Considering a signal of $x(u)$, the continuous Wavelet transform is expressed as: (Feng *et al.* (2013)).

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\, \psi \frac{(u-t)}{a}\, du \tag{5.1}$$

where $\psi(t)$ is the base wavelet and u and a are dilatation and translation factors. The wavelet $\psi(u-t)/a$ is derived by dilating and translating the wavelet basis $\psi(t)$. $1/\sqrt{a}$ is a normalization factor to maintain energy conservation and $a > 0$.

5.2.2 Convolutional Neural Network

Deep convolutional neural networks (CNNs) have recently attracted a lot of interest in various machine learning applications due to their ability to be trained on large-scale data with minimal preprocessing. CNNs are multi-stage neural networks consisting of one input layer, one or multiple convolutional layers, pooling layers, fully connected layers and one output layer. They have recently proven to be greatly successful in various machine learning tasks, such as regression and prediction. The convolution and pooling layers can appropriately characterize internal (intraclass) variations within large amounts of data. The CNN network performs very well in creating features from unprocessed signals by learning data-driven filters from the data to represent the practical information of the inputs instead of using traditional feature generation methods (Bouvrie (2006)).

The input signal is fed to a convolution layer where the feature maps are convolved with learnable kernels and construct the output feature map by the activation function. Each output map may combine convolutions with multiple input maps. The general formulation for a convolutional layer is (Bouvrie (2006)):

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{5.2}$$

Here, the feature map in the previous layer is X_i^{l-1} , f is a nonlinear activation function, M_j represents a selection of input maps, k is a square matrix with the size of kernels , l is the index for each convolution layer and b is an additive bias given to each output map.

The pooling layers are generally used after the convolutional layers to allow down-sampling or dimensionality reduction. Therefore, the dimension of the output maps will decrease but their number will remain the same as the number of input maps. As an example, max pooling is a function which uses the maximum value from each cluster of neurons at the prior layer to perform down-sampling (Ciregan *et al.* (2012)). In terms of formulation (Bouvrie (2006)):

$$X_j^l = f\left(\beta_j^l down(X_j^{l-1}) + b_j^l\right) \tag{5.3}$$

where $down()$ represents a sub-sampling function. Each output map has its own multiplicative bias β and an additive bias b.

As the last layer, fully connected layers are traditional multi-layer perceptions (MLPs) that (Ruck *et al.* (1990)) compute the desired outputs using an activation function from the previous layer output.

5.2.3 Transfer Learning

In traditional machine learning algorithms, sufficient labeled data must be available for training the model to ensure its accuracy. In cases where acquiring such data on a large scale is not possible, certain difficulties will be encountered when attempting to create a reliable model. Transfer learning is an emerging and promising solution that allows solving such difficulties by leveraging existing knowledge from the labeled data and avoiding excessive efforts for generating large labeled datasets for a similar task of interest. Transfer learning has the potential to greatly improve the performance of machine learning models for a given task using smaller data sizes.

In transfer learning, a machine learning model is trained with labeled data based on a specific feature space. This model could be used as an initial source model to create another model for a different feature space by partially retraining the machine learning algorithm on a limited and smaller size dataset which would mean reusing a previously learned model and knowledge.

Let us assume that $\mathscr{X}$ and $P(X)$ are a feature space and a probability distribution of domain D. $X = \{x_1, ..., x_i\}$ in $\mathscr{X}$ is a learning sample for space $\mathscr{X}$ with probability distribution of $P(X)$. Within a domain D =$\{\mathscr{X}, P(X)\}$, $f(.)$ is trained with the label space $(y_i \in Y, x_i \in X)$ with a learning task of $T = \{y, f(.)\}$ Assuming two domains D_s with learning task of T_s and D_t with learning task of T_t where $D_s \neq D_t$ and $T_s \neq T_t$, transfer learning will use the $f_s(.)$ to improve the $f_t(.)$.

Figure 5.1 represents the difference between the learning processes of traditional and transfer learning approaches. Based on the figure, traditional machine learning techniques try to learn each task from scratch, while transfer learning techniques aim to transfer the knowledge from some previous tasks to a target task when the latter has a smaller amount of high-quality training data.

5.3 Experimental Datasets

In this research, two different experimental milling datasets were used to demonstrate the transfer learning potential. The first dataset is the benchmark NASA-Ames and UC Berkeley milling dataset (Agogino & Goebel (2007)) obtained from milling operations with the Matsuura MC-510V machining center, with a 70 mm face mill and 6 KC710 inserts as cutter tools. The experiments were performed with the cutting speed set to 200 m/min for two different materials, cast iron and stainless steel J45, under varying cutting parameters (depth of cut, feed rate,

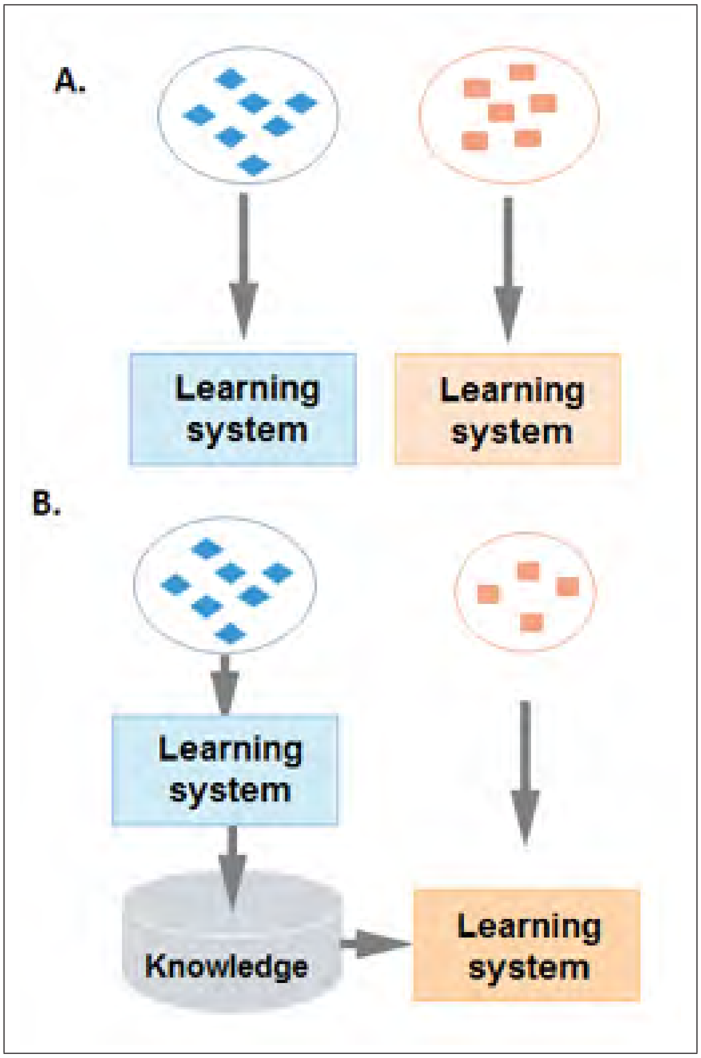

Figure 5.1 A. Traditional learning, B. Transfer learning

etc.). In total, 16 different cases of tests were performed, resulting in 146 experiments, and the tool flank wear was measured at different intervals. Flank wear is defined as the distance in μm from the cutting edge to the end of the abrasive wear on the flank face. Two accelerometers (model 720150, ENDEVCO) with a frequency range of up to 13 KHz were mounted on the table and spindle of the machine and fed to an RMS algorithm after amplification. A PHOENIX CONTACT UMK-SE11,25 cable connector was used for high speed data acquisition, with a maximal sampling rate of 100 KHz. Many experiments and different cutting parameters and materials were used to design an effective and diverse source model. The second dataset was the ÉTS Experimental Dataset, which the experiments were performed on the K2X10 Huron high speed CNC machine in the ÉTS LIPPS laboratory. The milling experiments were performed with a D2 high-speed tooling steel as the workpiece material with dimension of 200 x 54 x 4 mm. D2 steel is a known highly wear-resistant hard- to- cut material with a hardness of 60-62 HRC. The cutting tool was a Carbide Walter End Mill Protostar H50 Ultra tool with 6 teeth and a 50- degree helix angle. The experiments were performed under different cutting speeds (of 2500 rpm and 6000 rpm) and feed rates (0.12 mm/tooth and 0.05 mm/tooth), with 4 mm depth of cut, and the flank wear was measured at different intervals, resulting in 51 cases.

To measure the acceleration, two tri-axial accelerometers were mounted on the spindle and table of the machine with a sensitivity of 100 mV/g. Figure 5.3 represents this experimental setup. The robustness of the model against variable cutting parameters as well as different tools and machines will be verified using the aforementioned two datasets.

5.4 Proposed Methodology

5.4.1 Source Model Architecture

Figure 5.2 describes the methodology of this research. The source model is trained on the vibration signal of the NASA-Ames milling dataset from the accelerometers mounted on the machine spindle; where the RMS value of the vibration signals is reported by the data acquisition device. The signal is then filtered and fed to a signal processing module in which wavelet time frequency transformation is performed using the Morlet mother wavelet. This wavelet provides good resolution for both time and frequency domains. The output of the signal processing step is then fed to a feature extraction module where wavelet packet transform by PyWavelets (Wasilewski (2010)) is used to generate features by creating uniform frequency bands from the signal. Here, 4 decomposition levels are used and 16 uniform bands are generated, and all 16 are used to train the model. Since CNN is known for its ability to learn from

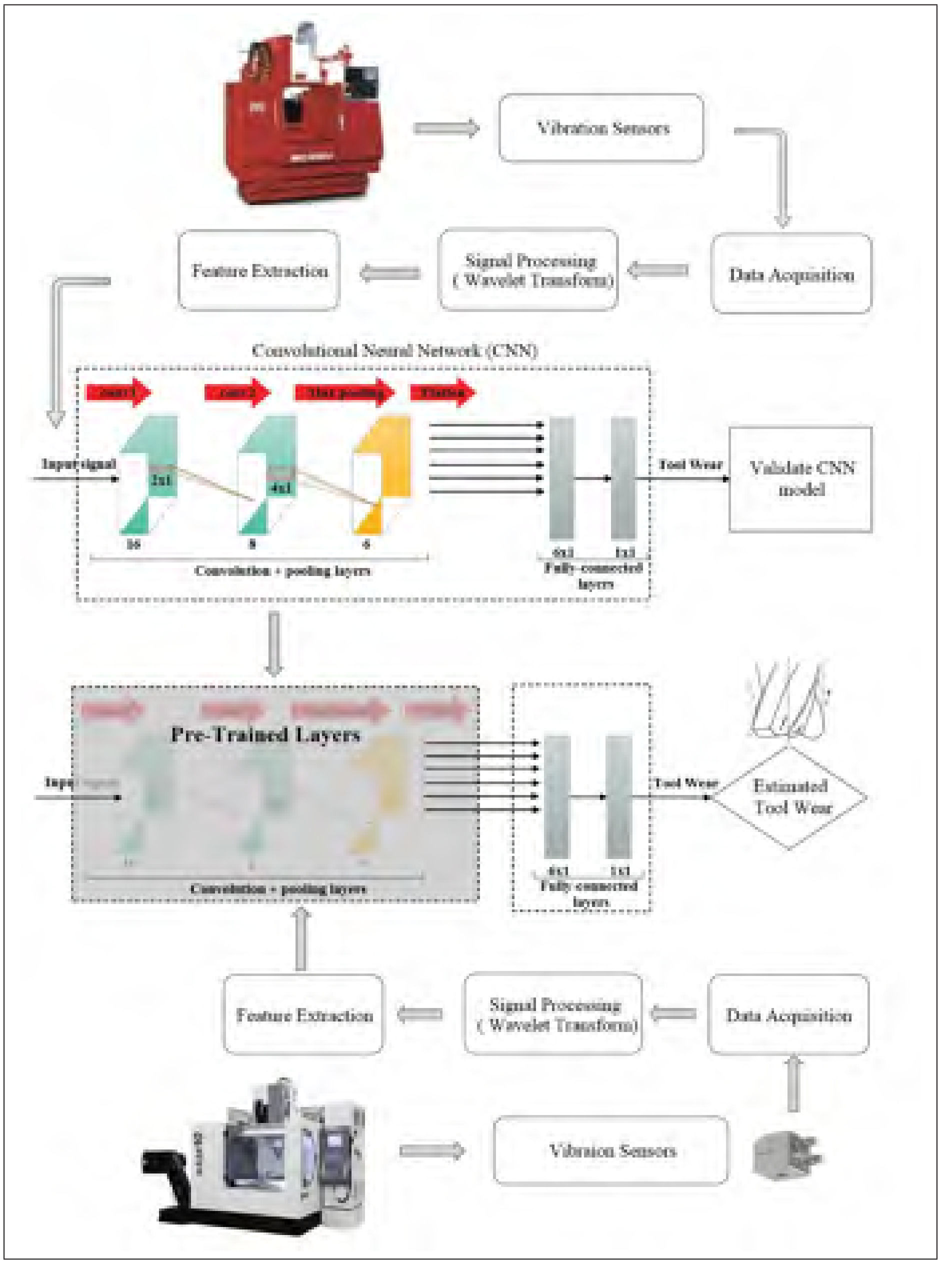

Figure 5.2 The monitoring system framework

minimally processed features, no further feature extraction is performed and the extracted features are directly fed to a CNN model.

In the next step, a regression CNN model is proposed to predict the value of flank wear. The proposed CNN model architecture includes the input layer where the 16 frequency bands are fed as the input to the two consecutive convolutional layers with dimensions of 16x13 and 10x8 respectively. The output of the two convolutional layers are fed to a max pooling layer with ReLu as the activation function to reduce the dimension and processing cost of the learning process. The output of this layer is then flattened and fed to a fully connected multi-layer perceptron neural network with two fully connected layers with the softmax activation function for prediction of tool wear value.

5.4.2 Target Model Architecture

The fully connected layers of the model described in previous section are retrained with the vibration signal of the ÉTS dataset to construct the target model. The ÉTS dataset is relatively smaller, and the goal is to demonstrate the efficiency of using a source model trained on a larger dataset to train a target model for the ÉTS experimental dataset.

In this model, as is the case in the last section, the RMS value of the signal is reported, filtered and fed to the same signal processing module and then, all 16 frequency bands are fed to the same CNN model. Training the convolutional layers of the model is the step requiring the highest amount of data and its role is to transform the signal to more informative features for fault detection. Since the general problem in both cases (source and target) are similar, we used the exact same training weights from the source model in the target model for the convolutional, pooling and flatten layers and converted the parameters of these layers to non trainable. This step reduces the training cost. The output from the flatten layer is fed to the fully connected layers. The parameters of these layers are re-tuned by the ÉTS dataset to adapt the source model to the target data space. Table 5.1 presents the layers, their training status and number of trainable parameters for both the source and target models.

5.5 Results

The source monitoring model is developed as per the methodology described in the previous section. Available signals are randomly divided into two categories, namely, training (80%)

Figure 5.3 Experimental set up

and testing(20%) datasets. The training dataset is used to train the source model in 600 epochs

Table 5.1 Model's layers information and their training status in source and target systems

Layer (type)	Output Shape	Number of Trainable Parameters	Source Model	Target Model
Convolutional 1	(1,13,16)	80	Trainable	Non-Trainable
Convolutional 2	(1,10,8)	520	Trainable	Non-Trainable
Max Pooling	(1,1,8)	0	Trainable	Non-Trainable
Flatten	8	0	Trainable	Non-Trainable
Fully Connected 1	8	72	Trainable	Trainable
Fully Connected 2	1	1	Trainable	Trainable

Table 5.2 Accuracy results of the target model

Algorithm	Source Model		Target Model	
	Average Accuracy %	RMSE	Average Accuracy %	RMSE
Transfer-Learning CNN	83.99	0.0236	80.05	0.0295
Support Vector Regression (SVR)	83.89	0.0250	65.15	0.0400
Bayesian Network	83.30	0.0244	61.96	0.0390

by minimizing the mean square error. Figure 5.4 represents the loss function value converging toward zero during the training process.

Once the source model is trained, it is used to re-train certain layers for the ÉTS dataset as discussed in the Methodology section. Vibration signals from the ÉTS dataset are also randomly divided into two categories, training (80%) and testing(20%) datasets. Figure 5.5 reports the loss values of the target model during 200 epochs of training method, which shows that it is converging toward zero.

Table 5.2 presents the results of tool wear estimation using the test dataset for both the source and target models. To evaluate and compare the accuracy and performance of the systems, the average accuracy in percentage (the differences between the predicted and actual tool wear value divided by the average of tool wear values) and RMSE are calculated as representative of the performance from the Scikit-learn machine learning performance analysis toolboxes. Moreover, the Support vector regression and Baysian network regression methods are implemented using the same dataset and processing steps as baseline for further comparison.

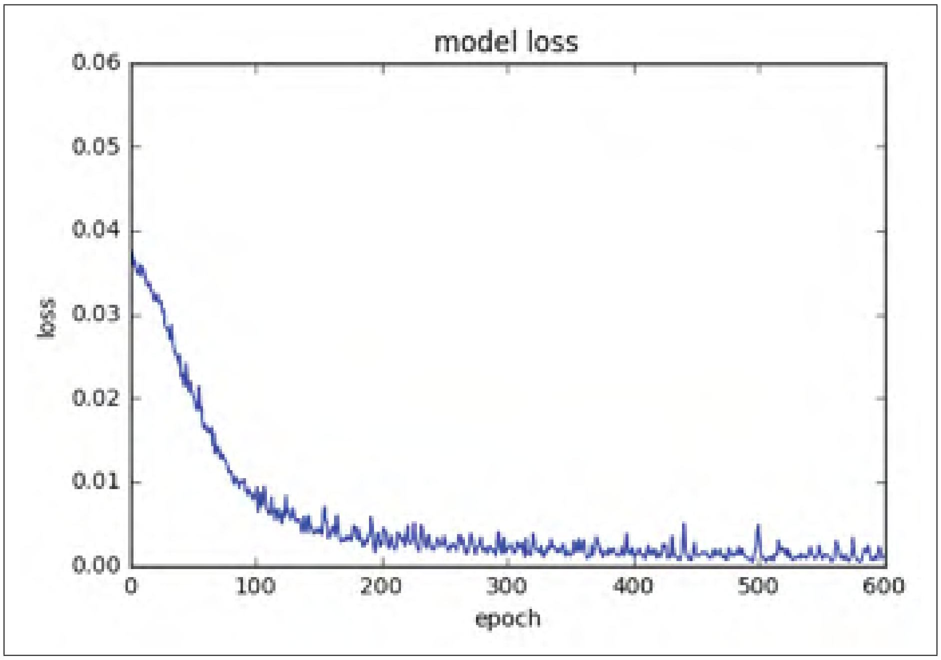

Figure 5.4 Loss function during training process of the source model

Based on Table 5.2, the accuracy of the base model is around 84% for all models and slightly higher in the CNN case. Deep learning algorithms require much larger datasets than the ETS or NASA Ames dataset to fully exploit their advantages, but they still provide high accuracies using smaller amount of data. Therefore, we can conclude that for applications with enough, yet limited available data, deep learning and powerful machine learning algorithms are both able to provide acceptable accuracy.

It is observed from Table 5.2 that the target model accuracy, which is the focus of this paper, is significantly (around 15%) higher for the transfer learning case. The modeldeveloped with transfer learning approach also produces a lower RMSE. The results suggest that an acceptable monitoring accuracy can be achieved even with smaller datasets by leveraging the transfer learning methodology. SVR and Bayesian models demonstrate relatively low accuracies which can be explained by the fact that the available training data is not adequate to allow them to learn the model behavior from scratch. Therefore, they are more prone to suffer from the under-fitting problem.

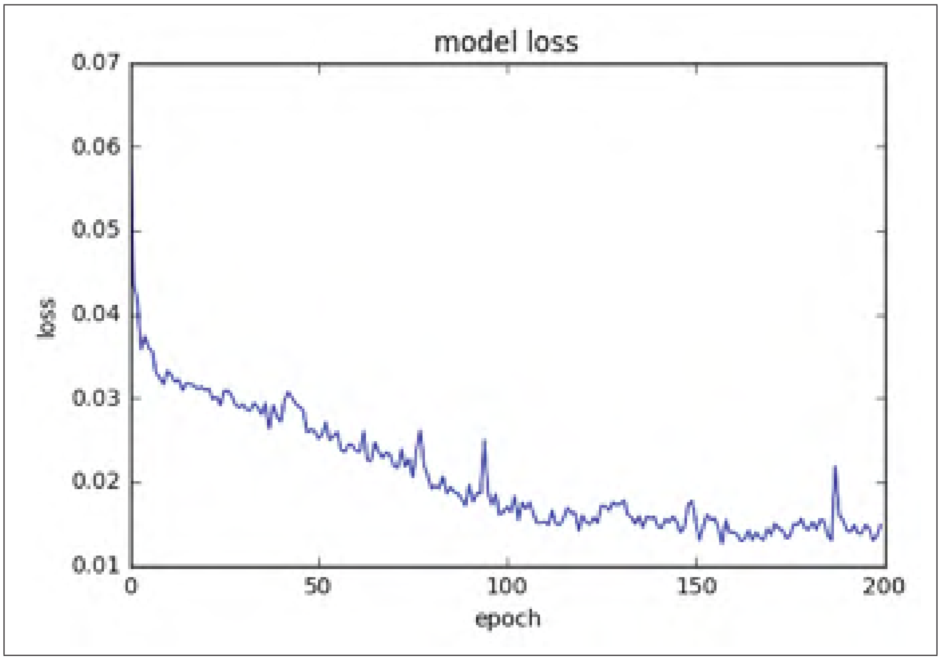

Figure 5.5 Loss function during training process of the target model

5.6 Conclusions

This research investigates the feasibility and performance of leveraging knowledge gained through the training of a monitoring system in a different machining center and operation conditions. The source model is trained based on the NASA-Ames dataset with various experiments. The model benefits from a Wavelet transform for signal processing and a deep CNN algorithm which relates the signals to tool wear. The target experiments are conducted on a different machining center, material and under different operating conditions with considerably lower available data than the NASA-Ames dataset. The deep transfer learning method is used to transfer the knowledge from the source model to the target problem. Certain layers of the source model are set to non-trainable and the last fully connected layers are retrained using target dataset. Finally the results are compared to two baseline machine learning algorithms.

Based on the results, in the source problem, the deep CNN model has a slightly higher accuracy than the other algorithms, but they all provide high accuracies of about 84%. It is observed that the source model accuracy is limited by available knowledge in the dataset rather than the training method. Moreover, for the target problem, the transfer learning model has a high accuracy of 80% despite the fact that it uses a relatively low number of experiments for training. This indicates that transfer learning performs very well event with a dataset that is not large. Two other algorithms, SVR and Baysian suffered from under-fitting because of the low amount of training data available. Consequently, they produce lower accuracies. Therefore, it is concluded that the transfer learning approach has the potential to overcome the high data requirement drawback in machine learning based approaches. A source model trained on benchmark data or similar datasets can significantly increase the accuracy of future models and reduce their data and experimental requirements. Therefore, this approach opens the door to the use of scalable and lower cost artificial intelligence based monitoring systems in the field of condition monitoring.

CHAPTER 6

SYNTHESIS

Tool wear is one of the most common faults during machining process which reduces productivity, finish surface quality and may cause unscheduled downtime. Automated monitoring of the tool condition and wear helps to overcome some of these problems by knowing beforehand the cutting tool condition and enables us to pro-actively maintain the tool condition to achieve higher productivity, maximize the tool usage and monitor the parts quality. Therefore, there is a high demand by manufacturing industry to continuously monitor tool condition. There are certain requirements for such system to be deployed for industrial applications. The main requirements include leveraging cost effective sensors which are practical for industrial environment, high accuracy of tool wear estimation with minimum false alarm and high scalability.

The core of current research is to design and develop an online tool condition monitoring system which is able to address most of those requirements. Different chapters of this thesis have studied these requirements and provided solutions to them. Therefore, combining the achievements and results of the all chapters of this thesis will enable us to design a robust online tool condition motoring system which was the main purpose of this study.

As the first step we should determine the fault indicators for our monitoring systems. In Chapter 1, we reviewed the previous work in literature to investigate what are the most common approaches and sensors used for tool condition monitoring, what per-processing and signal enhancement methodologies are performed in previous work for this application and what are the modeling techniques researchers explored in this domain. Based on this study we propose to use force, vibration and current signals as the fault indicators. A combination of vibration and current signals would be favorable since they are more practical in industry.

The acquired signals need to be processed to reduce the noises and extract relevant information to tool wear. Based on the study and analysis in Chapter 3, Time-frequency transformation method is proposed to perform this step in the final system since it focuses on both time domain and frequency domain characteristics of the signal simultaneously. Spectral subtraction is also favorable for noise reduction. It is an algorithm which is developed on top of the signals' time-frequency representation to reduce the steady state noise and focuses on the main subject of interest in a signal. It considers a signal as a combination of noise and clean information and by subtracting the noise from the original signal, clean signal can be achieved. Therefore, this set of algorithms conclude the signal processing step of the model.

The next step in designing the system is to identify the best artificial intelligence and machine learning methodologies to model the complex relationships between the extracted features and tool condition. We can leverage the findings in Chapter 4 which throughly examined conventional and deep learning methodologies to find the best algorithm. Convolutional neural networks is proposed in this part as a promising Deep Learning algorithm for tool condition monitoring. Deep Learning methodologies have recently showed a lot of potential in solving complex problems accurately in many applications. Especially with widespread availability of low cost sensors and advancements in the data storage and computation, it is key to leverage these algorithms to leverage the vast amount of available data. The unique characteristic of this method is its convolution layers which are able to transform the inputs to informative features using the automatically learned data-driven filters in the training step. Therefore, it is capable to exploit information from low processed data.

Scalability is another key requirement of the industrial condition monitoring systems which should be addressed in the final model. Machine learning methods work well under this assumption that the training and test data should come from the same feature space with the same distribution. Therefore, a model trained on a specific machine and for a certain task, may not be reusable for another machine or different tasks and conditions. To solve this problem and improve scalability of the intelligent tool condition monitoring systems we proposed transfer learning base on the findings in the Chapter 5 of the thesis. Transfer learning refers to re-use of the knowledge gained through solving a problem in a different yet similar problem. In transfer learning, a model is trained on a source task using a relatively large available dataset. Afterwards, the source model is used as the starting point to develop another model for a different target task with a limited and smaller size dataset. The second model is built by either fine tuning the source model or using certain layers of the source model and retraining some others. Therefore, Leveraging transfer learning and benchmark datasets in the field of tool condition monitoring will help us deliver more accurate monitoring systems with less available data using this approach. Furthermore, it would be more scalable to extend an excising model to work under other cutting operations, tasks and machines.

In summary, different chapters of this thesis addressed key requirements of design and development of a reliable tool condition monitoring system and by combining those, we were able to achieve an online reliable monitoring system. Force, vibration and current sensors are proposed as the fault indicators. Time-frequency transformation techniques and spectral subtraction are recommended for signal processing and enhancement to consider both time domain and frequency domain information of the signals simultaneously. Deep learning methodologies and in particular, CNNs is proposed to model complex relationships between extracted features and tool wear. Finally, a transfer learning framework is proposed to scale the monitoring systems and leverage previous data in enhancing future models.

CONCLUSIONS

In this thesis, tool condition monitoring under changing cutting parameters in milling process was investigated. Indirect approach is chosen for the monitoring due to its practicality. There are multiple factors hindering the development of a reliable monitoring system which this research aims to address. Tool wear is a major problem especially in machining of the hard to cut materials where the tool wears more quickly. In indirect monitoring methods, physical signals are used to indicate the tool wear. The signals in industrial environment are subjected to various noises and environmental factors which reduce the quality of them. Therefore, advanced signal processing techniques are required to clean the signals. Furthermore, the relationship between signals and tool wear are non-linear and highly complicated. Advanced algorithms are necessary to model this relationship. Finally, the scalability issue of monitoring systems needs to be addressed. Therefore, knowledge between monitoring systems with different machines and tasks should be transferable to reduce the effort for developing new monitoring systems. These are addressed in different chapters of the thesis.

Spectral subtraction method is investigated in the third chapter as an advanced signal processing technique. It has contributed highly in revealing the fault signature of the signal by removing the steady state part of the signal due to normal cutting and magnifying the remaining fault characteristics. Both in Chapter 3 and Chapter 4, this method demonstrated great performance in revealing fault characteristics in current signal.

In Chapter 4, deep CNN method is examined to model the complex relationships between extracted features and tool wear values. Some common machine learning algorithms such as Bayesian, KNN and SVR are also implemented as a baseline to compare the results. The results showed that CNN consistently outperforms other machine learning algorithms among all tested signals from multiple datasets. It is due to the CNN's convolution layers ability to alter the signal and extract discriminative features from the raw wavelet response.

The feasibility of leveraging the knowledge gained through training a tool wear monitoring system for milling in a different machining center and operation conditions is investigated in

Chapter 5. Deep transfer learning method is used for this purpose and the results are compared to two baseline machine learning algorithms. The first system is developed using a large benchmark dataset and CNNs network (source model) and the same model is used by slight retraining and tuning (target model) for tool wear monitoring on a new dataset. Based on the results, the transfer learning model can reach acceptable accuracy even with low number of experiments for training. This proves transfer learning performance when the dataset is not large. Therefore, transfer learning has potential to overcome the drawback of high data requirements in machine learning based approaches by opening the door to have scalable and lower cost artificial intelligence based monitoring systems in the field of condition monitoring.

Contributions

This research contributed in various aspects of the tool wear monitoring. As part of this research, a set of experiments are conducted for the machining of the D2 steel, a very hard to cut material. Machining of the hard to cut materials is a challenging task in industry as they may cause excessive tool wear or breakage during the machining. Force and vibration signal are captured during experiments to indirectly report tool faults.

The next step of designing a condition monitoring system, is signal processing. Wavelet packet transform is leveraged as an established method for fault diagnosis. In the case of current signals, spectral subtraction is introduced to this field as a signature extractor due to the scaler nature of current signals. spectral subtraction was able to increase around 5% the accuracy of the current signal based model.

Machine learning techniques are used widely in the literature for fault modeling and automation of the monitoring. A comparative study between the main conventional machine learning methods is conducted to evaluate their performance in the new dataset for the machining of the hard to cut materials. Among the conventional methods, decision trees regressors method showed highest accuracy for fault estimation.

Powerful capabilities of deep learning methods draw attention of researchers in many different fields to solve complex challenges in machine learning domain. Widespread availability of low-cost sensors and data also has paved the way for deep learning to generate accurate data-driven models with minimal pre-processing of the data. This research investigates the application of these methods in the field of machining monitoring. CNNs are throughly examined with various different sensors (force, vibration and current) and datasets and the results show it outperforms consistently the baseline conventional machine learning methods. Recurrent neural networks is also examined to capture the sequential and time sensitive aspect of the data.

Transfer learning is also introduced and examined in this application. Machine learning algorithms have high volume of data requirements and most of them work well only when training and test data come from the same feature space with the same distribution. Transfer learning tackles this issue by leveraging the knowledge gained while training a system, and adopting it and making it usable for another system. This is investigated and shown as another contribution of this research and the results show with limited volume of the data, acceptable monitoring models can be achieved when leveraging the knowledge of other previous systems.

The significant contributions made in this thesis are resulted in the following publications in a chronological order:

- Aghazadeh, Fatemeh, M. Thomas, and A. Tahan. "Effect of varying cutting parameters on the accuracy of tool wear monitoring in milling process.", Canadian Machinery Vibration Association Annual General Meeting and Seminar (2016), Toronto, Canada (Aghazadeh *et al.* (2016)).

- Aghazadeh, Fatemeh, Antoine Tahan, and Marc Thomas. "Tool condition monitoring using spectral subtraction algorithm and artificial intelligence methods in milling process." International Journal of Mechanical Engineering and Robotics Research 7.1 (2018): 30-34 (Aghazadeh *et al.* (2018b)).

- Aghazadeh, Fatemeh, Antoine Tahan, and Marc Thomas. "Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process." The International Journal of Advanced Manufacturing Technology 98.9-12 (2018): 3217-3227 (Aghazadeh *et al.* (2018a)).

- Aghazadeh, Fatemeh, Antoine S. Tahan, and Marc Thomas. "Tool condition monitoring method in milling process using wavelet transform and long short-term memory." Surveillance, Vishno and AVE conferences, INSA-Lyon, Université de Lyon,2019, Lyon, France (Aghazadeh *et al.* (2019)).

- Aghazadeh, Fatemeh, Antoine S. Tahan, and Marc Thomas. "Tool condition monitoring method in milling process using deep transfer learning" submitted to The International Journal of Advanced Manufacturing Technology, 2019.

RECOMMENDATIONS

Recommendations

Although this research tried to address some of the main challenges in design and development of an intelligent tool condition monitoring systems, there is a lot of room to enhance these systems in different aspects. Some of the recommendations for future work are summarized as follows:

1) The dataset in this research is acquired as an academic dataset with limited number of samples. Accessing direct industrial data at large scale is highly beneficial to further validating the intelligent tool condition monitoring. Different aspect of the research such as resistance to noise, scalability and applicability can be validated at the industry scale using bigger datasets.

2) In this research, we validated the models using multiple signals independently (Force, vibration, current) and sensor fusion and combining the data were considered out of the scope of it. Considering widespread availability of low cost sensors, study of sensor fusion techniques is a beneficial next step in this domain to enhance accuracy and reliability of the monitoring systems. A comparative study on combining data at different maturity levels (raw signals, extracted features or decision level) as well as examining the possibility of leveraging deep learning methods for sensor fusion is key for this field. Moreover, the relation between tool wear and surface quality can be investigated.

3) More in-depth research is necessary in the signal processing step to enhance signal quality and focus on the fault signatures. This research employed wavelet packet transform for directional signals and spectral subtraction for the scalar signals. Research on applying methods similar to spectral subtraction but tailored for vector type signals can enhance their quality. Moreover, kinematic variables in rotating machinery are periodic with respect to some rotation angles. Therefore, analysis of signals using cyclostationary techniques and in angular domain is another approach which can help to reveal discriminative features from signals.

4) Further research on the Deep Learning methodology is highly beneficial in improving state of the art methods for condition monitoring. First angle can be using directly raw signals as the input of these methods as they are famous for being able to exploit information from raw signals. This will reduce the cost of feature engineering and pre-processing, however, it requires subnational amount of data at much larger scales to be able to exploit data-driven information from raw signals. Moreover, deep learning is a live field with continuous advancement in the methods and architectures. Latest methods of deep learning should be evaluated in this application with potential improvement in accuracy, reliability or applicability.

5) This research presented a transfer learning approach for easier scaling of the monitoring system and being able to transfer knowledge between systems. While we explored one method of transfer learning in an application, further analysis should explore other methods of transfer learning with different architecture. For example, in this study, both source and target problems were supervised (with clear fault labels). Another approach is using a large dataset of the sensors' signal to learn a intermediary representation of the signals with unlabeled data and then applying the signal representations learned from the unsupervised learning to a supervised fault estimation task. Moreover, in this study, we froze the first layers for the target model. Instead of freezing the first layers, they can be retrained as well, but with the very good initialization from source model compared to training them from scratch. Finally, CNN based transfer learning is investigated in this research, however, transfer learning method can be applied to other algorithms such as RNNs, LSTMs, etc.

6) This research utilized an average size benchmark dataset for building its source model for transfer learning and leveraged the knowledge gained in this model for a smaller target dataset. However, utilizing a large scale dataset with more variability of test cases, material and machines is highly beneficial in generalizing the source model. This introduces a huge potential among researches to enhance the tool condition monitoring in a collaborative way. A long term recommendation would be preparing a baseline source model leveraging multiple datasets in the industry for tool condition monitoring. Therefore, researches and professionals can lever-

age that as the starting point of their research and build on top of that using transfer learning to reduce data requirements and leveraging directly previous work and datasets in this domain.

7) In this research, transfer learning is explored for a similar task (Tool condition monitoring) and a similar application (milling operation) but with different machining centers, materials and cutting parameters. The future work can investigate knowledge transfer in a broader scale such as exploring the possibility of knowledge transfer between different type of faults, operations and applications. For example, it can be investigated how a source model in the milling operation is applicable as a baseline for a turning, drilling or even bearing fault monitoring systems.

BIBLIOGRAPHY

High Speed Steel - D2 Steel - D2 Technical Data. Accessed: 2019-06-30.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H. & Penn, G. (2012). Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*, pp. 4277–4280.

Abellan-Nebot, J. V. & Subirón, F. R. (2010). A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology*, 47(1-4), 237–257.

Aghazadeh, F., Thomas, M. & Tahan, A. (2016). Effect of varying cutting parameters on the accuracy of tool wear monitoring in milling process. *Canadian Machinery Vibration Association Annual General Meeting and Seminar*.

Aghazadeh, F., Tahan, A. & Thomas, M. (2018a). Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. *The International Journal of Advanced Manufacturing Technology*, 98(9-12), 3217–3227.

Aghazadeh, F., Tahan, A. & Thomas, M. (2018b). Tool condition monitoring using spectral subtraction algorithm and artificial intelligence methods in milling process. *International Journal of Mechanical Engineering and Robotics Research*, 7(1), 30–34.

Aghazadeh, F., Tahan, A. S. & Thomas, M. (2019). Tool condition monitoring method in milling process using wavelet transform and long short-term memory. *Surveillance, Vishno and AVE conferences, INSA-Lyon, Université de Lyon, Lyon, France*.

Aghdam, B., Vahdati, M. & Sadeghi, M. (2015). Vibration-based estimation of tool major flank wear in a turning process using ARMA models. *The International Journal of Advanced Manufacturing Technology*, 76(9-12), 1631–1642.

Agogino, A. & Goebel, K. (2007). Mill data set. *BEST lab, UC Berkeley. NASA Ames Prognostics Data Repository," NASA Ames, Moffett Field, CA, http://ti. arc. nasa. gov/project/prognostic-data-repository*.

Aliustaoglu, C., Ertunc, H. M. & Ocak, H. (2009). Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system. *Mechanical Systems and Signal Processing*, 23(2), 539–546.

Altintas, Y. (1992). Prediction of cutting forces and tool breakage in milling from feed drive current measurements. *Journal of engineering for Industry*, 114(4), 386–392.

Ammouri, A. & Hamade, R. (2014). Current rise criterion: a process-independent method for tool-condition monitoring and prognostics. *The International Journal of Advanced Manufacturing Technology*, 72(1-4), 509–519.

Barros, J. & Diego, R. I. (2006). Application of the wavelet-packet transform to the estimation of harmonic groups in current and voltage waveforms. *IEEE Transactions on Power Delivery*, 21(1), 533–535.

Baziotis, C., Athanasiou, N., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S. & Potamianos, A. (2018). NTUA-SLP at SemEval-2018 Task 1: predicting affective content in tweets with deep attentive RNNs and transfer learning. *arXiv preprint arXiv:1804.06658*.

Bengio, Y., Courville, A. & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798–1828.

Benkedjouh, T., Medjaher, K., Zerhouni, N. & Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, 26(2), 213–223.

Benkedjouh, T., Zerhouni, N. & Rechak, S. (2018). Tool wear condition monitoring based on continuous wavelet transform and blind source separation. *The International Journal of Advanced Manufacturing Technology*, 97(9-12), 3311–3323.

Bhuiyan, M. & Choudhury, I. (2014). 13.22—Review of sensor applications in tool condition monitoring in machining. *Comp. Mater. Process*, 13, 539–569.

Bhuiyan, M., Choudhury, I., Dahari, M., Nukman, Y. & Dawal, S. (2016). Application of acoustic emission sensor to investigate the frequency of tool wear and plastic deformation in tool condition monitoring. *Measurement*, 92, 208–217.

Bishop, C. M. et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Bodin, P. & Villemoes, L. F. (1997). Spectral subtraction in the time-frequency domain using wavelet packets. *Speech Coding For Telecommunications Proceeding, 1997, 1997 IEEE Workshop on*, pp. 47–48.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, 27(2), 113–120.

Bouvrie, J. (2006). Notes on convolutional neural networks.

Bre, F., Gimenez, J. M. & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158, 1429–1441.

Chen, Z., Li, C. & Sanchez, R.-V. (2015). Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration*, 2015.

Cho, S., Binsaeid, S. & Asfour, S. (2010). Design of multisensor fusion-based tool condition monitoring system in end milling. *The International Journal of Advanced Manufacturing Technology*, 46(5-8), 681-694.

Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. *URL: https://keras.io/k*, 7(8), T1.

Choqueuse, V., Benbouzid, M. E. H. et al. (2013). Current frequency spectral subtraction and its contribution to induction machines' bearings condition monitoring. *IEEE Transactions on Energy Conversion*, 28(1), 135–144.

Ciregan, D., Meier, U. & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 3642–3649.

Dechter, R. (1986). *Learning while searching in constraint-satisfaction problems*. University of California, Computer Science Department, Cognitive Systems ....

Delio, T., Tlusty, J. & Smith, S. (1992). Use of audio signals for chatter detection and control. *Journal of engineering for industry*, 114(2), 146–157.

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, 3.

DirectIndustry. (2016). The online Industrial Exhibition: sensor - automation - motor - pump - handling - packaging [online photo].

Djurović, I., Sejdić, E. & Jiang, J. (2008). Frequency-based window width optimization for S-transform. *AEU-International Journal of Electronics and Communications*, 62(4), 245–250.

Dutta, S., Pal, S. K. & Sen, R. (2016). On-machine tool prediction of flank wear from machined surface images using texture analyses and support vector regression. *Precision Engineering*, 43, 34–42.

Elgargni, M., Al-Habaibeh, A. & Lotfi, A. (2015). Cutting tool tracking and recognition based on infrared and visual imaging systems using principal component analysis (PCA) and discrete wavelet transform (DWT) combined with neural networks. *The International Journal of Advanced Manufacturing Technology*, 77(9-12), 1965–1978.

Feng, Z., Liang, M. & Chu, F. (2013). Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mechanical Systems and Signal Processing*, 38(1), 165–205.

Freyer, B. H., Heyns, P. S. & Theron, N. J. (2014). Comparing orthogonal force and unidirectional strain component processing for tool condition monitoring. *Journal of Intelligent Manufacturing*, 25(3), 473–487.

Funahashi, K.-i. & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6), 801–806.

Gangadhar, N., Kumar, H., Narendranath, S. & Sugumaran, V. (2014). Fault diagnosis of single point cutting tool through vibration signal using decision tree algorithm. *Procedia Materials Science*, 5, 1434–1441.

Girardin, F., Rémond, D. & Rigal, J.-F. (2010). Tool wear detection in milling—An original approach with a non-dedicated sensor. *Mechanical Systems and Signal Processing*, 24(6), 1907–1920.

Gowid, S., Dixon, R. & Ghani, S. (2015). A novel robust automated FFT-based segmentation and features selection algorithm for acoustic emission condition based monitoring systems. *Applied Acoustics*, 88, 66–74.

Guo, L., Lei, Y., Xing, S., Yan, T. & Li, N. (2018). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*.

Harun, M., Ghazali, M. & Yusoff, A. (2017). Analysis of tri-axial force and vibration sensors for detection of failure criterion in deep twist drilling process. *The International Journal of Advanced Manufacturing Technology*, 89(9-12), 3535–3545.

Huang, P., Li, J., Sun, J. & Zhou, J. (2013). Vibration analysis in milling titanium alloy based on signal processing of cutting force. *The International Journal of Advanced Manufacturing Technology*, 64(5-8), 613–621.

Hurmuzlu, Y. & Nwokah, O. D. (2016). *The mechanical systems design handbook: modeling, measurement, and control*. CRC Press.

Iglesias-Martinez, M. E., de Cordoba, P. F., Antonino-Daviu, J. & Conejero, J. A. (2019). Detection of Nonadjacent Rotor Faults in Induction Motors via Spectral Subtraction and Autocorrelation of Stray Flux Signals. *IEEE Transactions on Industry Applications*.

Imaouchen, Y., Alkama, R. & Thomas, M. (2015). Bearing fault detection using motor current signal analysis based on wavelet packet decomposition and Hilbert envelope. *MATEC Web of Conferences*, 20, 03002.

ISO8688-2. (1989). *ISO 8688-2 : Tool Life Testing in Milling. Part 2 : End Milling*. ISO. Consulted at https://books.google.ca/books?id=ymgiMwEACAAJ.

Jarrett, K., Kavukcuoglu, K., Ranzato, M. & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? *2009 IEEE 12th international conference on computer vision*, pp. 2146–2153.

Jáuregui, J. C., Reséndiz, J. R., Thenozhi, S., Szalay, T., Jacsó, Á. & Takács, M. (2018). Frequency and Time-Frequency Analysis of Cutting Force and Vibration Signals for Tool Condition Monitoring. *IEEE Access*, 6, 6400–6410.

Jia, F., Lei, Y., Lin, J., Zhou, X. & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72, 303–315.

Jin, X., Zhao, M., Chow, T. W. & Pecht, M. (2014). Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Transactions on Industrial Electronics*, 61(5), 2441–2451.

Jing, L., Zhao, M., Li, P. & Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, 111, 1–10.

Kedadouche, M., Thomas, M. & Tahan, A. (2014). Monitoring machines by using a hybrid method combining MED, EMD, and TKEO. *Advances in Acoustics and Vibration*, 2014.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Kosaraju, S., Anne, V. G. & Popuri, B. B. (2013). Online tool condition monitoring in turning titanium (grade 5) using acoustic emission: modeling. *The International Journal of Advanced Manufacturing Technology*, 67(5-8), 1947–1954.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pp. 1097–1105.

Kuljanic, E., Totis, G. & Sortino, M. (2009). Development of an intelligent multisensor chatter detection system in milling. *Mechanical Systems and Signal Processing*, 23(5), 1704–1718.

Lamraoui, M., Thomas, M. & El Badaoui, M. (2014a). Cyclostationarity approach for monitoring chatter and tool wear in high speed milling. *Mechanical Systems and Signal Processing*, 44(1-2), 177–198.

Lamraoui, M., Thomas, M., El Badaoui, M. & Girardin, F. (2014b). Indicators for monitoring chatter in milling based on instantaneous angular speeds. *Mechanical Systems and Signal Processing*, 44(1-2), 72–85.

Lamraoui, M., Barakat, M., Thomas, M. & Badaoui, M. E. (2015). Chatter detection in milling machines by neural network classification and feature selection. *Journal of Vibration and Control*, 21(7), 1251–1266.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E. & Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, pp. 396–404.

Li, N., Chen, Y., Kong, D. & Tan, S. (2017). Force-based tool condition monitoring for turning process using v-support vector regression. *The International Journal of Advanced Manufacturing Technology*, 91(1-4), 351–361.

Lin, J. & Qu, L. (2000). Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration*, 234(1), 135–148.

Lin, X., Zhou, B. & Zhu, L. (2017). Sequential spindle current-based tool condition monitoring with support vector classifier for milling process. *The International Journal of Advanced Manufacturing Technology*, 1–10.

Lipton, Z. C., Kale, D. C., Elkan, C. & Wetzel, R. (2015). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.

Luo, B., Wang, H., Liu, H., Li, B. & Peng, F. (2019). Early fault detection of machine tools based on deep learning and dynamic identification. *IEEE Transactions on Industrial Electronics*, 66(1), 509–518.

Madhusudana, C., Kumar, H. & Narendranath, S. (2017). Face milling tool condition monitoring using sound signal. *International Journal of System Assurance Engineering and Management*, 8(2), 1643–1653.

Malhi, A. & Gao, R. X. (2004). PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, 53(6), 1517–1525.

Marinescu, I. & Axinte, D. A. (2008). A critical analysis of effectiveness of acoustic emission signals to detect tool and workpiece malfunctions in milling operations. *International Journal of Machine Tools and Manufacture*, 48(10), 1148–1160.

McClelland, J. L., Rumelhart, D. E., Group, P. R. et al. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, 2.

Moore, R. & Lopes, J. (1999). Paper templates. *TEMPLATE'06, 1st International Conference on Template Production*.

Muralidharan, V. & Sugumaran, V. (2012). A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Applied Soft Computing*, 12(8), 2023–2029.

Nee, A. Y. C. (2015). *Handbook of manufacturing engineering and technology*. Springer.

Ogedengbe, T. I., Heinemann, R., Hinduja, S. et al. (2011). Feasibility of tool condition monitoring on micro-milling using current signals. *AU JT 14 (3)*, 161–172.

Pai, S., Nagabhushana, T. & Rao, R. B. (2012). Tool condition monitoring using acoustic emission, surface roughness and growing cell structures neural network. *Machining Science and Technology*, 16(4), 653–676.

Pan, S. J. & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345–1359.

Pan, S. J. & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345–1359.

Pandiyan, V., Caesarendra, W., Tjahjowidodo, T. & Tan, H. H. (2018). In-process tool condition monitoring in compliant abrasive belt grinding process using support vector machine and genetic algorithm. *Journal of Manufacturing Processes*, 31, 199–213.

Patra, K., Jha, A., Szalay, T., Ranjan, J. & Monostori, L. (2017). Artificial neural network based tool condition monitoring in micro mechanical peck drilling using thrust force signals. *Precision Engineering*, 48, 279–291.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.

Prakash, M., Ravisankar, P. & Kanthababu, M. (2014). Acoustic Emission Signal Analysis for Tool Condition Monitoring in Microendmilling of Aluminium Alloy. *Advanced Materials Research*, 984, 25–30.

Proteau, A., Tahan, A. & Thomas, M. (2019). Specific cutting energy: a physical measurement for representing tool wear. *The International Journal of Advanced Manufacturing Technology*, 1–10.

Quintana, G. & Ciurana, J. (2011). Chatter in machining processes: A review. *International Journal of Machine Tools and Manufacture*, 51(5), 363–376.

Rad, J. S., Hosseini, E., Zhang, Y. & Chen, C. (2013). Online tool wear monitoring and estimation using power signals and S-transform. *2013 Conference on Control and Fault-Tolerant Systems (SysTol)*, pp. 234–238.

Rad, J. S., Zhang, Y., Aghazadeh, F. & Chen, Z. C. (2014a). A study on tool wear monitoring using time-frequency transformation techniques. *Innovative Design and Manufacturing (ICIDM), Proceedings of the 2014 International Conference on*, pp. 342–347.

Rad, J. S., Zhang, Y. & Chen, C. (2014b). A Novel Local Time-Frequency Domain Feature Extraction Method for Tool Condition Monitoring Using S-Transform and Genetic Algorithm. *IFAC Proceedings Volumes*, 47(3), 3516–3521.

Rangwala, S. & Dornfeld, D. (1990). Sensor integration using neural networks for intelligent tool condition monitoring. *Journal of engineering for industry*, 112(3), 219–228.

Rehorn, A. G., Jiang, J. & Orban, P. E. (2005). State-of-the-art methods and results in tool condition monitoring: a review. *The International Journal of Advanced Manufacturing Technology*, 26(7), 693–710.

Rehorn, A. G., Sejdić, E. & Jiang, J. (2006). Fault diagnosis in machine tools using selective regional correlation. *Mechanical Systems and Signal Processing*, 20(5), 1221 - 1238.

Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E. & Suter, B. W. (1990). The multi-layer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, 1(4), 296–298.

Salakhutdinov, R. & Hinton, G. (2009). Deep boltzmann machines. *Artificial intelligence and statistics*, pp. 448–455.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85–117.

Segreto, T., Simeone, A. & Teti, R. (2013). Multiple sensor monitoring in nickel alloy turning for tool wear assessment via sensor fusion. *Procedia CIRP*, 12, 85–90.

Shi, D. & Gindy, N. N. (2007). Tool wear predictive model based on least squares support vector machines. *Mechanical Systems and Signal Processing*, 21(4), 1799 - 1814. doi: https://doi.org/10.1016/j.ymssp.2006.07.016.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285–1298.

Siddhpura, A. & Paurobally, R. (2013). A review of flank wear prediction methods for tool condition monitoring in a turning process. *The International Journal of Advanced Manufacturing Technology*, 65(1-4), 371–393.

Siddhpura, M. & Paurobally, R. (2012). A review of chatter vibration research in turning. *International Journal of Machine tools and manufacture*, 61, 27–47.

Snr, D. E. D. (2000). Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. *International Journal of Machine Tools and Manufacture*, 40(8), 1073–1098.

Solatian, P., Abbasi, S. H. & Shabaninia, F. (2012). Simulation study of flow control based on PID ANFIS controller for non-linear process plants. *American journal of intelligent systems*, 2(5), 104–110.

Stavropoulos, P., Chantzis, D., Doukas, C., Papacharalampopoulos, A. & Chryssolouris, G. (2013). Monitoring and control of manufacturing processes: A review. *Procedia CIRP*, 8, 421–425.

Stockwell, R. G. (2007). A basis for efficient representation of the S-transform. *Digital Signal Processing*, 17(1), 371–393.

Stockwell, R. G., Mansinha, L. & Lowe, R. (1996). Localization of the complex spectrum: the S transform. *Signal Processing, IEEE Transactions on*, 44(4), 998–1001.

Tansel, I. N. (1992). Modelling 3-D cutting dynamics with neural networks. *International Journal of Machine Tools and Manufacture*, 32(6), 829–853.

TechniksUSA. (2019). CNC tool holders, Collet Chucks, End Mill Holders, ER Collets, and Accessories [online photo].

Tobon-Mejia, D., Medjaher, K. & Zerhouni, N. (2012). CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. *Mechanical Systems and Signal Processing*, 28, 167–182.

Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103.

Wang, G., Guo, Z. & Yang, Y. (2013). Force sensor based online tool wear monitoring using distributed Gaussian ARTMAP network. *Sensors and Actuators A: Physical*, 192, 111–118.

Wang, G., Yang, Y. & Li, Z. (2014). Force sensor based tool condition monitoring using a heterogeneous ensemble learning model. *Sensors*, 14(11), 21588–21602.

Wang, G., Zhang, Y., Liu, C., Xie, Q. & Xu, Y. (2019). A new tool wear monitoring method based on multi-scale PCA. *Journal of Intelligent Manufacturing*, 30(1), 113–122.

Wang, J., Xie, J., Zhao, R., Mao, K. & Zhang, L. (2016). A new probabilistic kernel factor analysis for multisensory data fusion: application to tool condition monitoring. *IEEE Transactions on Instrumentation and Measurement*, 65(11), 2527–2537.

Wang, P. & Lee, C.-M. (2019). Fault Diagnosis of a Helical Gearbox Based on an Adaptive Empirical Wavelet Transform in Combination with a Spectral Subtraction Method. *Applied Sciences*, 9(8), 1696.

Wang, W., Hong, G.-S., Wong, Y. & Zhu, K. (2007). Sensor fusion for online tool condition monitoring in milling. *International Journal of Production Research*, 45(21), 5095–5116.

Wasilewski, F. (2010). Pywavelets: Discrete wavelet transform in python.

Yan, J. & Lee, J. (2005). Degradation assessment and fault modes classification using logistic regression. *Journal of manufacturing Science and Engineering*, 127(4), 912–914.

Yang, J., Zhang, D., Frangi, A. F. & Yang, J.-y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1), 131–137.

Yen, C.-L., Lu, M.-C. & Chen, J.-L. (2013). Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting. *Mechanical Systems and Signal Processing*, 34(1-2), 353–366.

Yesilyurt, I. & Ozturk, H. (2007). Tool condition monitoring in milling using vibration analysis. *International journal of production research*, 45(4), 1013–1028.

Zhang, C. L., Yue, X., Jiang, Y. T. & Zheng, W. (2010). A hybrid approach of ANN and HMM for cutting chatter monitoring. *Advanced Materials Research*, 97, 3225–3232.

Zhang, S., Li, J. & Wang, Y. (2012). Tool life and cutting forces in end milling Inconel 718 under dry and minimum quantity cooling lubrication cutting conditions. *Journal of Cleaner Production*, 32, 81–87.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. & Gao, R. X. (2016). Deep learning and its applications to machine health monitoring: A survey. *arXiv preprint arXiv:1612.07640*.

Zhao, R., Yan, R., Wang, J. & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17(2), 273.

Zhou, Y. & Xue, W. (2018). Review of tool condition monitoring methods in milling processes. *The International Journal of Advanced Manufacturing Technology*, 96(5-8), 2509–2523.

Zhu, K. & Vogel-Heuser, B. (2014). Sparse representation and its applications in micro-milling condition monitoring: noise separation and tool condition monitoring. *The International Journal of Advanced Manufacturing Technology*, 70(1-4), 185–199.

Zhu, K. & Zhang, Y. (2019). A generic tool wear model and its application to force modeling and wear monitoring in high speed milling. *Mechanical Systems and Signal Processing*, 115, 147–161.

Zuo, W.-M., Wang, K.-Q. & Zhang, D. (2005). Assembled matrix distance metric for 2DPCA-based face and palmprint recognition. *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, 8, 4870–4875.