APPLICATIONS OF DATA MINING TOOLS IN MACHINING

THESIS SUBMITTED BY

SHRUTI SUDHAKAR DANDGE

DOCTOR OF PHILOSOPHY (ENGINEERING)

DEPARTMENT OF PRODUCTION ENGINEERING

FACULTY COUNCIL OF ENGINEERING & TECHNOLOGY

JADAVPUR UNIVERSITY

KOLKATA, INDIA

2025

JADAVPUR UNIVERSITY
KOLKATA-700 032, INDIA

Index No. 74/19/E
Registration No. 1021916004

TITLE OF THE Ph.D. (Engg.) THESIS:

APPLICATIONS OF DATA MINING TOOLS IN MACHINING

NAME, DESIGNATION & INSTITUTION OF THE SUPERVISOR:

Dr. SHANKAR CHAKRABORTY
Professor, Department of Production Engineering,
Jadavpur University,
Kolkata –700032, India.

LIST OF PUBLICATIONS

Title of Thesis	**:**	**APPLICATIONS OF DATA MINING TOOLS IN MACHINING**
Name, Designation and Institute of the Supervisor:	**:**	**Dr. Shankar Chakraborty** Professor Department of Production Engineering Jadavpur University, Kolkata, India

List of Publications:

a) In International Journal

Sl. No.	Title of the paper	Name of the Journal
1	Decision tree-based parametric analysis of a CNC turning process	Scientia Iranica, 2021; 28(6):3653-3674.
2	A data mining approach for analysis of a wire electrical discharge machining process	Management and Production Engineering Review, 2021;12(3):116-128.

b) In Proceedings of Conference/Seminars

Sl. No.	Title of the paper	Name of the Conference Proceedings
1	Selection of machining parameters in ultrasonic machining process using CART algorithm	In: Venkata Rao R., Taler J. (eds.) Advanced Engineering Optimization through Intelligent Techniques, Springer, Singapore, 2019, 949.

c) List of Patents: Nil

d) List of presentations in National/International Conference/Seminar

Sl. No.	Title of the paper	Name of the Conference
1	Selection of machining parameters in ultrasonic machining process using CART algorithm	International Conference on Advanced Engineering Optimization through Intelligent Techniques, 3-5 August 2018, Surat, Gujarat, India.
2	A comparative performance analysis of predictive models for an electrical discharge machining process	International Conference on Precision, Meso, Micro and Nano Engineering, 13-15 December 2024, NIT Calicut, India.

STATEMENT OF ORIGINALITY

I, **Shruti Sudhakar Dandge** registered on 30.04.2019 do hereby declare that this thesis entitled "**APPLICATIONS OF DATA MINING TOOLS IN MACHINING**" contains literature survey and original research work done by the undersigned candidate as part of Doctoral studies.

All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is **9%**.

S. S. Dandge

Signature of the Candidate

Date: 21/08/2025

Certified by Supervisor

Shankar Chakraborty 21/08/25

(Dr. Shankar Chakraborty)

Thesis Advisor

Department of Production Engineering

JADAVPUR UNIVERSITY

Kolkata-700032

Dr. Shankar Chakraborty
Professor
Department of Production Engg.
Jadavpur University
Kolkata – 700 032

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled "**APPLICATIONS OF DATA MINING TOOLS IN MACHINING**" submitted by **Miss. Shruti Sudhakar Dandge**, who got her name registered on 30.04.2019 for the award of Ph. D. (Engg.) degree of Jadavpur University is absolutely based upon her own work under the supervision of **Dr. Shankar Chakraborty** and that neither her thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

Shankar Chakraborty 21/04/25

(Dr. Shankar Chakraborty)

Thesis Advisor

Department of Production Engineering

JADAVPUR UNIVERSITY

Kolkata-700032

Dr. Shankar Chakraborty
Professor
Department of Production Engg.
Jadavpur University
Kolkata – 700 032

ACKNOWLEDGEMENT

I take this opportunity to extend my sincere gratitude and appreciation to all those who helped me in completing the dissertation.

First and foremost, I would like to express my deep sense of gratitude towards my honourable thesis advisor Dr. Shankar Chakraborty, Professor, Department of Production Engineering, Jadavpur University for his generous and ardous guidance, encouragement and continuous support throughout my research work. His enthusiasm, integral view on research and mission for providing high quality work has helped me in improving the quality of my research work.

I would also like to offer my sincere gratitude to Government Polytechnic, Murtijapur, Maharashtra for permitting me to pursue my Ph. D. research work. I am also thankful to the respected Principal, Government Polytechnic, Arvi for his valuable support whenever required. I also take this opportunity to thank all the faculty members and staffs of my workplace for constantly supporting and encouraging me throughout the research work. I am equally grateful to the Head of Department of Production Engineering, Jadavpur University for providing departmental facilities during this research work.

My special words of thanks to my fellow colleagues, like Sunny Diyaley, Vidyapati Kumar, Shubham Agrawal, Sudeep Santra, Devendra Pendhokare, Gajanan Padmane, Baneshwar Sarkar and Saikat Chatterjee for their constant cooperation, useful assistance and support during this research work.

The support, best wishes and blessings of my parents and my family have been a constant source of inspiration and encouragement throughout this research work. I express my gratitude to my dearest mother Mandakini Dandge for her kind support and giving me strength throughout my life.

I also acknowledge the support of all those people, who directly or indirectly helped me in completing the research work.

Lastly, I bow my head to the Almighty God for providing me enough courage, strength and patience for successful completion of this research work.

SS Dandge

VITA

The author, Miss. Shruti Sudhakar Dandge, is currently working as a Lecturer in Department of Mechanical Engineering, Government Polytechnic, Arvi, Maharashtra, India. She has almost 14 years of academic experience in engineering and polytechnic institutions. During this tenure, she has taught many subjects, like Engineering Thermodynamics, Fluid Power, Power Plant Engineering, Engineering Graphics, Mechanical Drawing, I. C. Engines etc.

She has obtained her Bachelor's degree in Mechanical Engineering in 2006 from V.Y.W.S. College of Engineering, Badnera, Maharashtra. She has then obtained her Master's degree in 2009 with specialization in Thermal Engineering from Mechanical Engineering Department, Government College of Engineering, Amravati, Maharashtra. She has attended numerous workshops, and faculty development programs in the area of technology and management. She has qualified Graduate Aptitude Test in Engineering (GATE) examination in the year 2007 conducted by Department of Higher Education, Ministry of Human Resources Development, India.

Her research interests are in the area of applications of different data mining techniques in machining environment. She has guided several undergraduate and diploma project theses. She has presented several papers in international and national conferences as well as published some papers in international journals of high repute.

PREFACE

Since the last few decades, the entire scenario of manufacturing has changed significantly shifting from the conventional to computer numerical control, robotics and flexible systems. This has become necessary due to increase in cost of labour, reduction in lead time and rising demand for higher product quality. In order to meet the ever increasing requirements of low cost high quality products, enhanced production rate, higher dimensional accuracy and lower surface finish, the conventional machining operations are being continuously substituted by the computer-based and automated technologies. The achievable precision and accuracy through the adoption of automated operations cannot be obtained by the traditional material removal processes.

In manufacturing industries, machining is one of major operations performed to remove unwanted material from the workpiece surface to attain the desired shape of the final product/component while fulfilling the customers' needs. Machining thus involves shaping of a component through removal of material. This can be achieved by the use of a tool, having material harder than the component being shaped, causing material removal by shear deformation in the form of chips. The performance of any of the machining operations can usually be characterized by the combination of its various input parameters and outputs (responses).

In advanced machining processes, these input parameters need to be more precisely controlled so as to achieve the target response values, reduce machining time and cost, minimize tool wear, impart more flexibility, generate complicated shapes, provide higher repeatability, attain high dimensional tolerance, reduce power consumption etc. It has been observed that there exist complex relationships between the input parameters and responses in the machining operations. Understanding these relationships and proposing techniques for parametric optimization in improving performance of the machining processes play pivotal roles in the present-day manufacturing environment.

Some of the machining processes related to the present research work are mentioned as below:

a) Computer numerical control (CNC) turning process,
b) Wire electrical discharge machining (WEDM) process,
c) Ultrasonic machining (USM) process, and
d) Electrical discharge machining (EDM) process

Optimization of a machining process has already been identified as a complex problem due to involvement of several different and often contradictory objectives, like maximization of material removal rate and minimization of surface roughness, maximization of machining efficiency and minimization of power consumption etc. It is the primary objective of any optimization tool to identify the ideal values of various machining parameters so as to achieve better process performance. Usually, in manufacturing industries, the concerned machinists select the most suitable settings of different input parameters based on their knowledge and expertise. Sometimes, machining data handbooks have also been consulted to meet those requirements. However, the proposed machining parameters are far from their optimal values which may hinder achievement of the goal of better performance. Because of availability of voluminous data with rapid advancement of machining technology, attainment of the optimal machining performance cannot now be possible only with the deployment of conventional optimization techniques. Therefore, the need for more advanced tools is ardently felt to fulfil the above-mentioned objective.

Since the last few decades, rapid advancements in computational facilities and information technology have enabled the human beings to move beyond manual, tiring and labourious practices for quick, straightforward and automated analysis of data. In this direction, data mining (DM) is a powerful tool with immense applicability and potentiality to turn raw data into useful information. DM, occasionally known as ‘knowledge discovery in databases’, is the procedure of drilling through the data to unveil hidden patterns/connections and anticipate future trends. The fundamental task of DM algorithms is to identify those attribute values effective for predicting other attributes. DM involves use of various advanced tools and techniques to refine and extract valuable knowledge from large databases. This process aims to categorize, characterize and envisage the mined data, identifying interesting patterns that facilitate valuable decision-making. Unlike popular statistical and predictive models, DM can uncover patterns even when data sources are heterogeneous and differently distributed. The core of DM lies in applying various mathematical tools, including machine learning (ML), cluster analysis, regression analysis and neural networks (NNs). The decision tree (DT) analysis can be defined as a set of practices to estimate and show the extant relations between a dependent variable and a class of independent variables. It is based on consecutive partitioning algorithm to continuously split data to constitute homogeneous subsets, developing a hierarchical tree consisting of decision rules convenient for data anticipation or classification. The main classification methods for DT induction are Classification and Regression Tree (CART) and Chi-squared Automatic Interaction Detection (CHAID) algorithms. As the name proposes, association rules are straightforward 'If-Then' statements used to analyze frequently occurring patterns in a dataset or uncover inherent relationships between independent and dependent variables. Predictive modeling is used to forecast future outcomes based on previously known data. It involves proposing a mathematical model that captures relationships within the data to forecast future predictions or behaviours. The artificial neural networks (ANNs) are designed and developed based on human brain, which consists of a large number of interconnected elements known as neurons.

These neurons collaborate to process information, recognize patterns and relationships within data, and make necessary predictions.

Applications of DM in the manufacturing domain are still under-utilized and infrequently used since modern manufacturing systems and processes are very complex, dynamic and complicated. From a large number of available machining processes, some of them, like CNC turning, WEDM, USM, EDM etc. are considered in this research work for parametric analysis and prediction using various DM tools. For having enhanced performance of these machining processes, investigation of the influences of various input parameters on the responses and determination of the optimal combinations of the considered process parameters are very important. It is always necessary to understand the relationships between the process parameters and responses using appropriate mathematical models with some assumptions. DM helps incomprehensively uncovering those relationships beneficial for the machining processes.

With the increasing automation and rapid development of the computational intelligence, knowledge induction has received significant attention, specially in the domain of non-traditional machining (NTM) processes, to build competitive advantages. Knowledge induction from data has now become extremely important in NTM processes so as to enhance productivity, understand process mechanisms and improve future process performance. DM thus provides new iterative and interactive techniques and tools for processing large volume of data in order to explore valuable and understandable patterns hidden in the dataset. In the literature, various statistical methods and artificial intelligence (AI) techniques are considered as the suitable approaches for developing predictive models for understanding the complex material removal behaviour in various machining processes. But, availability of huge machining and manufacturing related data in digital form has also accelerated the application of DM tools to aid the process engineers in identifying the tentative settings of various input parameters to obtain the desired response values.

Thus, in this research work, applications of DM tools in different machining processes are proposed. Popular DM tools in the form of DT, clustering, association rules, ANN, extreme gradient boosting (XGB) and predictive modeling techniques are applied to analyze the experimental data of different machining processes, i.e. CNC turning, WEDM, USM and EDM. The organization of this present research work is as follows: Chapter 1 introduces the concept of a machining process considering its working principle and classification, history of DM, along with the need of application of DM in machining (or subtractive manufacturing) processes. Review of the past literature showing various attempts in DM applications in the broad domain of manufacturing with special emphasis on machining, and objectives and scope of the present research work are also presented in this chapter. Chapter 2 illustrates introduction to the multidisciplinary field of DM. It discusses the evolutionary path of information technology which has led to the need for DM, and importance of its application. This chapter presents the general classification of various DM tools based on the type of applications that is targeted. It also overviews methods for classification, including DT induction, association rule mining to analyze patterns, and discover relationships between the dependent and independent variables in the datasets. It includes advanced predictive models, such as XGB, Kriging (KRG), radial basis function (RBF) and gene expression programming (GEP) that analyze data patterns to determine future outcomes. Additionally, advanced methods, such as ANN model to determine relationships within the data and make necessary predictions, and basic concepts of various statistical error metrics, are also presented in this chapter. Chapter 3 deals with parametric analysis of three machining processes, i.e. CNC turning, WEDM and USM using DT. This chapter first investigates use of DM tools, specifically DTs, to determine the optimal settings for various input parameters during a CNC turning operation. Two DT algorithms, i.e. CART and CHAID, are employed to examine the effects of the CNC turning parameters on the responses. The 'If-Then' rules extracted from both the DTs are used to investigate effects of the input parameters on the considered responses. The relative performance of both the algorithms is also compared with respect to solution accuracy and prediction risk. The CART algorithm outperforms CHAID with respect to higher overall classification accuracy and lower prediction risk. In the second case, parametric analysis of a WEDM process is also studied by applying the non-parametric DT algorithms. Two DT-based classification algorithms, i.e. CART and CHAID, are again employed here to analyze the effects of six WEDM parameters on four responses and identify the optimal parametric combination for achieving the desired outcomes. Additionally, a comparative analysis of the classification performance of CART and CHAID algorithms indicates that CART is superior, providing higher overall classification accuracy and reduced prediction risk. In the third case, CART analysis is employed to investigate the effects of various machining parameters on the responses during an USM operation. Chapter 4 describes applications of four powerful predictive modeling techniques for an EDM process. To envisage the performance characteristics of the considered EDM process, predictive models are developed using four techniques, like XGB, KRG, RBF and GEP. A comparison is carried out between the experimental response values and predicted values obtained from the developed models. The prediction performance of all the investigated modeling techniques is finally compared using various statistical error metrics. It is observed that XGB, KRG and RBF have better accuracy based on those statistical metrics and overall datasets in predicting the considered response values during the EDM process. Chapter 5 presents application of another important DM tool, i.e. ANN in an EDM process. Various NN models, e.g. feed forward NN (FFNN), convolutional NN (CNN), recurrent NN (RNN) and long short term memory-based (LSTM) recurrent NN are employed here for prediction of the responses during an EDM operation. In order to further validate prediction performance of the four NN models, values of different statistical error metrics are computed to identify the most suitable model for prediction of the process responses. A detailed comparative analysis of the prediction performance is also presented to decide the best NN model. It is observed that all the developed NN models show strong correlations between the experimental and predicted values, thus indicating significance of those models for prediction of the process performance. The LSTM and RNN demonstrate good predictive ability with the envisaged values within the range of actual values for most of the responses under consideration. Lastly, Chapter 6 consolidates the outcomes of this research work found from various investigations along with the limitations and future scope.

The researcher has thereby attempted to justify the present research work by exploring various DM tools for parametric analysis, and predictive modeling for envisaging performance of the considered machining processes, along with the relevant comparative analysis. Thus, these DM tools can be effectively applied to other traditional and NTM processes to investigate contributions of the input parameters on the responses and identify the most suitable parametric combinations for exploring their fullest machining potential. Based on the derived results, it can be clearly revealed that applications of DM tools can derive effective and efficient solutions for distinct machining problems. The applicability of DM for parametric analysis, and performance prediction of other traditional and NTM processes may also be explored as a future scope of the present research work.

TABLE OF CONTENTS

ABBREVIATIONS

AI	Artificial intelligence
ANFIS	Adaptive neuro-fuzzy interface system
ANN	Artificial neural network
ANOVA	Analysis of variance
BRANN	Bayesian regularized artificial neural network
CART	Classification and regression tree
CAT	Cat boost regression
CC	Cobalt content
CF	Cutting force
CHAID	Chi-squared automatic interaction detection
CIR	Circularity
CNC	Computer numerical control
CNN	Convolutional neural network
CS	Cutting speed
CYL	Cylindricity
DD	Dimensional deviation
DM	Data mining
DOC	Depth of cut
DT	Decision tree
E	Machining environment
ECM	Electrochemical machining
EDM	Electrical discharge machining
ELM	Extreme learning machine
ELSVM	Extreme learning support vector machine
EWR	Electrode wear rate
FCGR	Fatigue crack growth rate
FFNN	Feed forward neural network
FL	Fuzzy logic
FR	Feed rate
GA	Genetic algorithm
GBR	Gradient boosting regression
GEP	Gene expression programming
GMA	Gas metal arc
GP	Genetic programming
GPAR	Gaussian process autoregressive regression
GPR	Gaussian process regressor
GRA	Grey relational analysis
GRC	Grey relational coefficient
GRG	Grey relational grade

GRNN	General regression neural network
GS	Grit size
HKELM	Hybrid kernel extreme learning machine
HSTR	High strength and temperature resistance
I	Current
I_P	Peak current
KDD	Knowledge discovery in data mining
KNN	K-nearest neighbor
KRG	Kriging
LSTM	Long short term memory
LTB	Larger-the-better
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MCDA	Multi-criteria decision analysis
ML	Machine learning
MR	Machining rate
MRR	Material removal rate
MSE	Mean square error
NC	Numerical control
NN	Neural network
NR	Nose radius
NTM	Non-traditional machining
OC	Overcut
PC	Power consumption
PR	Power rating
RBF	Radial basis function
ReLU	Rectified linear unit
RF	Random forest
RMI	Root cause machine identifier
RMSE	Root mean squared error
RNN	Recurrent neural network
RSM	Response surface methodology
RST	Rough set theory
SBDL	Statistical batch-based decision tree learning
SR	Surface roughness
STB	Smaller-the-better
SV	Spark gap voltage
SVM	Support vector machine
SVR	Support vector regressor
TC	Taper cut
TL	Tool life

TM	Traditional machining
T_{off}	Pulse-off time
T_{on}	Pulse-on time
TP	Tool profile
TW	Thickness of workpiece
TWR	Tool wear rate
USM	Ultrasonic machining
V	Voltage
WEDM	Wire-electrical discharge machining
WF	Wire feed rate
WT	Wire tension
WWR	Wire wear ratio
XGB	Extreme gradient boosting
XGBR	Extreme gradient boosting regression

Dedicated to my parents (Aai and Baba)

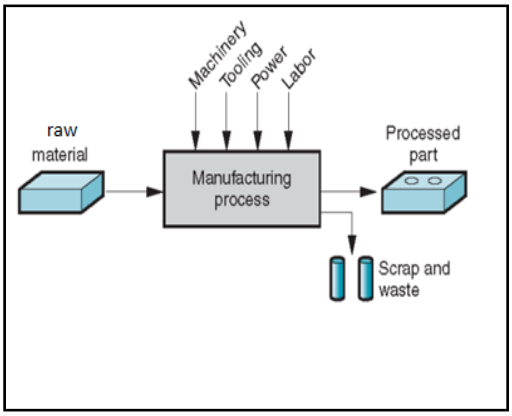
INTRODUCTION

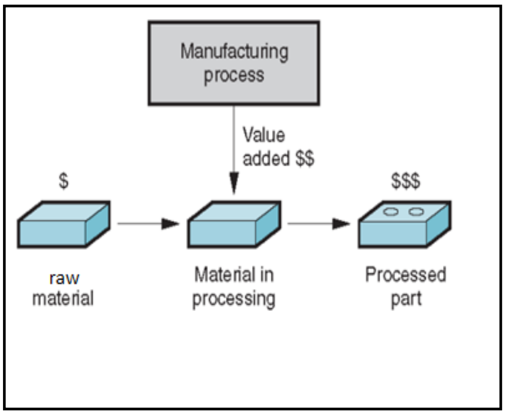
1. INTRODUCTION

1.1 An introduction to machining processes

Manufacturing constitutes the economic backbone of any industrialized nation and therefore, the manufacturing practices have become crucially important in the industrial environment for producing goods/products that serve humanity. Manufacturing is the value-adding procedure that converts raw materials into finished products, satisfying the customers' requirements. It is derived from the Latin word 'manufactus' which translates to 'made by hand'. In the modern context, it involves producing goods or products from raw materials through variety of procedures employing hand tools, machines and computers on much larger scale and with higher precision. Modern manufacturing is generally accomplished with automation and computer-controlled machines. In general, manufacturing is the economical term which involves number of linked activities and operations, including marketing, material selection, process planning, inventory control, quality assurance, and product design and development [1]. Therefore, the study of manufacturing identifies the parameters that can most efficiently boost production and improve accuracy.

There are two aspects in which manufacturing can be defined, i.e. technological and economical. Technologically, manufacturing is the field which deals with application of physical and chemical processes to modify the properties, geometry and appearance of the raw materials to transform into finished parts or products, also includes assembly of multiple components to make the final product. Manufacturing is the process that uses machinery, tools and equipment, power and labour, as depicted in Figure 1.1(a). It is always completed as a series of operations which aims at bringing the raw material closer to the final desired stage. From the economical point of view, manufacturing is the conversion of raw materials into final products possessing greater value by performing one or more processes and/or assembly operations, as shown in Figure 1.1(b).





a) b)

Figure 1.1 Definition of manufacturing: a) technological aspect and b) economical aspect [2]

An important factor to consider is that manufacturing 'adds value' to the raw materials by altering their shapes and properties or by combining them with other similarly modified materials [2]. The material would thereby undergo different manufacturing operations to increase its value. Thus, manufacturing can be interpreted as the study of various processes required to produce parts/products and assemble them through the required machines or mechanisms. Therefore, it is evident that manufacturing is not a single activity, but comprises various components interacting with each other dynamically to provide the necessary physical inputs at appropriate stages to attain the desired outputs. It thus requires large number of activities, few independent and rest mostly interrelated which jointly contribute towards economic and qualitative acceptable manufacturing of the desired products in minimum possible time and at reduced cost [3]. Thus, manufacturing, in its most comprehensive sense, involves transforming raw materials into usable products. This process includes designing the product, choosing the raw materials and determining the sequence of operations for production.

Most of the manufactured products require machining at some point during their production processes. Machining is the broad term used to describe removal of the undesired material from a workpiece to attain the finished product with desired shape, size and surface quality. Products manufactured by other primary shaping processes, such as casting, forging, bending, powder metallurgy etc. often require further operations before they are ready for application. The parts/components undergoing these operations are roughly finished products received through the primary shaping processes, requiring one or more machining operations in order to obtain the desired shape and dimensional precision. Machining is one of the most important stages of the manufacturing process in which a sharp cutting tool removes the excess material from the workpiece in the form of chips to achieve the desired geometry. This can be achieved manually or by using machines called machine tools. The process of chip formation is influenced by the relative motion between the cutting tool and the workpiece, obtained with the help of a machine tool. This relative motion is achieved by the combination of rotary and translatory motions of either tool, workpiece or both. While machining, the type of the surface required to be generated is determined by the tool's shape and the path it traverses through the material. These processes require one or more machine tools, different types of cutting tool, work holding devices, measuring equipment, testing devices etc. to reach the desired dimensional accuracy and surface finish. During machining, three types of chip are mainly formed, e.g. continuous chips, discontinuous chips and built-up-edge chips [4]. Continuous chips typically form while machining ductile materials, like wrought iron, aluminium, nickel, mild steel and copper, with high cutting speeds and rake angles. They are generally better at producing smoother surface finish, but are not necessarily always desirable, particularly while using computer controlled machine tools since they tend to get tangled around the tool holder, fixtures and the workpiece. This issue can be resolved with the help of chip breakers and adjusting the machining conditions, such as feed rate (FR), depth of cut (DOC) and cutting speed (CS), as well as using appropriate cutting fluids. Discontinuous chips are formed while machining brittle materials, like ceramics etc. because they cannot withstand high shear strains involved during the cutting operations. Very low or very high CS, higher DOC and low rake angle are generally responsible for formation of discontinuous chips during machining. Built-up-edge chips consist of multiple layers of workpiece material that progressively accumulate on the tool tip. Built-up-edge is a commonly observed phenomenon which adversely affects surface quality of the workpiece. The probability of formation of built-up-edge can be minimized by increasing CS, reducing DOC, increasing rake angle or using a sharp tool.

The machining processes range from coarse cleaning of casting to highly precise automated operations requiring tight tolerances. Modern advancements in the manufacturing domain necessitate transfer of more skill to the machines along with considerable reduction in human involvement. Parts/components of varying dimensions, forms, and increasing level of accuracy and complexities are constantly required to satisfy the demands of wide range of consumers and industries. Traditional methods with mechanical machining were first established to tackle these machining requirements. Traditional machining (TM) includes some basic machining operations, such as turning, drilling, milling, boring, grinding etc. where removal of material is mainly achieved by cutting and abrasion [5].

In TM, a tool, harder than the workpiece material, is employed which penetrates with a certain DOC into the workpice to be machined. During machining by cutting, the relative motion between the tool and the workpiece removes material in the form of chips to produce the parts/components of the required shape and dimensions. These TM processes have evolved into more sophisticated techniques as the related areas of manufacturing technology and material science continue to develop for tools and workpieces. To fulfil the needs of the emerging advanced industries in the field of aerospace, automobile, defence etc., the researchers in the areas of material science and metallurgy are developing materials which posses higher strength, hardness, toughness, high-strength temperature-resistance (HSTR) and other diverse properties. There is also a need for development of more advanced cutting tools in today’s competitive manufacturing environment which is loaded with mechanization, automation and computer-integrated technology. With significant advancements in materials offering increased strength, hardness, toughness etc., their processing using TM techniques, such as turning, milling and drilling, has really become challenging. It is a commonly known fact that for materials with high strength and hardness, using TM is not only feasible but also uneconomical. Tool materials which are sufficiently hard and strong to cut materials, like titanium, stainless steel, HSTR alloys etc. are not readily available. Production of complex shapes with high level of surface finish, minimum tolerance and higher production rate for those materials is still more difficult which is restricting the use of TM processes. Under these circumstances, to machine such materials with higher degree of accuracy and economy, advanced material removal processes have been developed without using traditional wedge-shaped tools. They are known as non-traditional machining (NTM) processes. Various forms of energy, such as thermal, electrical, mechanical, chemical or suitable combination of them are utilized in NTM processes to remove material from the workpiece to explore advantages and minimize potential disadvantages found in the traditional material removal techniques [6]. In the present-day manufacturing industries, use of these NTM processes is becoming increasingly unavoidable, because of their capability of generating complex shape geometries with tight tolerance, better surface finish, increased production rate etc. These processes have still become important when applications based on precision machining and ultra-precision machining are required. They have also their distinct characteristic features. Therefore, selection of an appropriate machining process to fulfil the product requirements becomes very crucial.

1.2 Classification of the machining processes

In the manufacturing domain, there are multiple methods that can be adopted to produce different components from the given raw material. Machining is the most prevalent method for shaping metals in manufacturing industries. Machining process can be broadly classified into two categories, i.e. TM and NTM processes, as depicted in Figure 1.2. In TM processes, sharp edge cutting tools are employed for material removal by means of physical contact with the workpiece to generate the required shape. Here, material removal mainly takes place due to cutting and abrasion.

Machining by cutting: These machining systems consist of a tool, workpiece and a machine tool which controls the relative motion between the tool and the workpiece. During machining by cutting, tool penetrates into the workpiece with certain DOC and removes the material by shearing action in the form of chips.

Machining by abrasion: In abrasive machining, material is removed by the action of hard abrasive (non-metallic) particles against the workpiece.

With the advent of new engineering materials and advancement of technology, cutting and abrasion machining processes have become very difficult because they employ tools that are harder than the workpiece. In addition, higher tool wear, excessive heat generation at the tool-workpiece interface, difficulty of machining harder materials, inaccessible or uneconomical to generate complex geometries, poor surface quality etc. are some of the limitations associated with TM processes. To overcome such limitations, NTM processes have been developed in the modern manufacturing scenario. They can be applied successfully where the workpiece surface is too hard, involving generation of complex shapes with both external and internal profiles and micro-features with high degree of accuracy and surface quality.

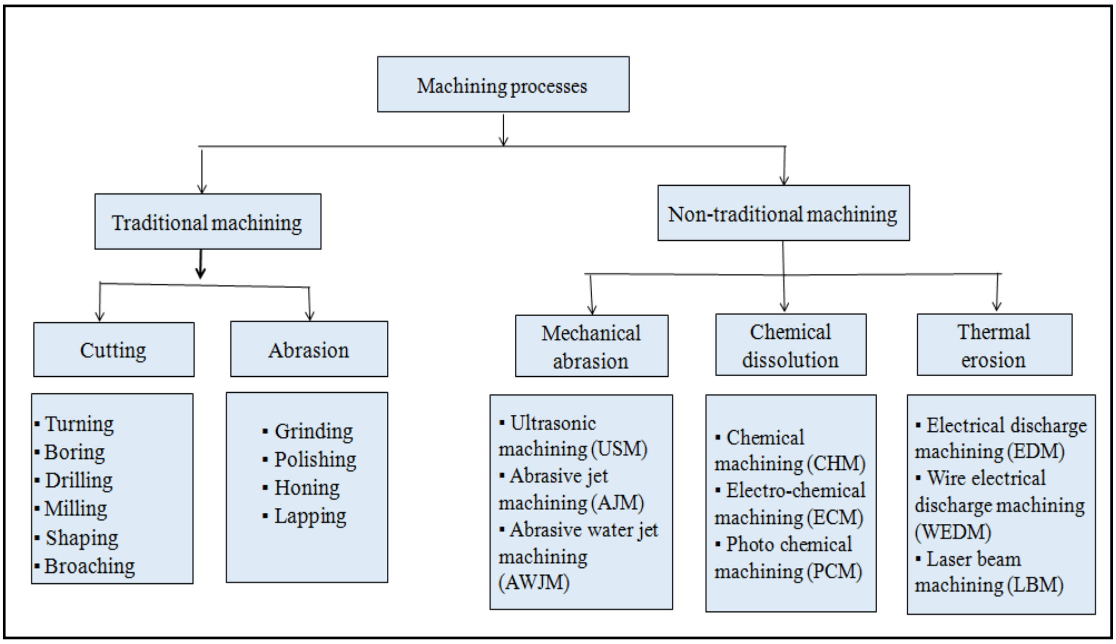


Figure 1.2 Classification of the machining processes

NTM processes utilize different types of energy, such as mechanical, thermal, electrical, chemical or a synergic combination of them to process the material, and convert it into final product having the desired shape, size and surface finish. This energy is utilized in a controlled manner using proper machinery and tooling, while human effort is dedicated to operate the setup, oversee the process, and loading and unloading of the components before and after each operational cycle. Hence, it becomes necessary to categorize these NTM processes to understand the basic mechanism of material removal.

Machining by erosion: In erosion machining, removal of successive surface layers of the work material takes place as a result of dissolution or melting and vaporization of the material being machined.

Mechanical abrasion: In this machining process, kinetic energy of abrasives, water jets as well as ions are utilized for removal of material.

Chemical and electro-chemical erosion: Chemical machining involves in controlled dissolution of the work material by application of strong reactive chemicals. On the other hand, in electro-chemical erosion, material is removed due to electro-chemical action, i.e. ion displacement where the workpiece is made anode and the tool is treated as cathode.

Thermal erosion: Machining by thermal erosion involves application of very intensive heat onto the workpiece surface to remove material by melting and evaporating small areas of the workpiece.

The working principles of different machining processes, specifically considered in the present research work, are introduced here-in-under.

Turning

Turning is a crucial machining operation where a single-point cutting tool removes excess material from a rotating workpiece to generate a cylindrical shape [7]. The cutting tool is fed linearly along the axis of rotation. This process is performed on a lathe machine, which provides the necessary power to rotate the workpiece at a specific speed, and controls the FR and DOC for the cutting tool. Figure 1.3 represents a typical turning operation of a cylindrical workpiece with initial diameter (D_i), final diameter (D_f) and length (L). CS, FR and DOC are considered as the three important input parameters during the turning operation. Since the turning operation employs a cutting tool, high forces and temperatures involved have detrimental effects on the tool. Additionally, the goal of the turning operation is to achieve lower surface roughness (SR) of the workpiece. Therefore, tool life (TL) and SR are the two important measures for evaluation of the cutting performance. During machining, path followed by the tool during its feed motion onto the work surface generates additional work shapes which include straight turning, taper turning, contour turning, facing, thread cutting, boring etc.

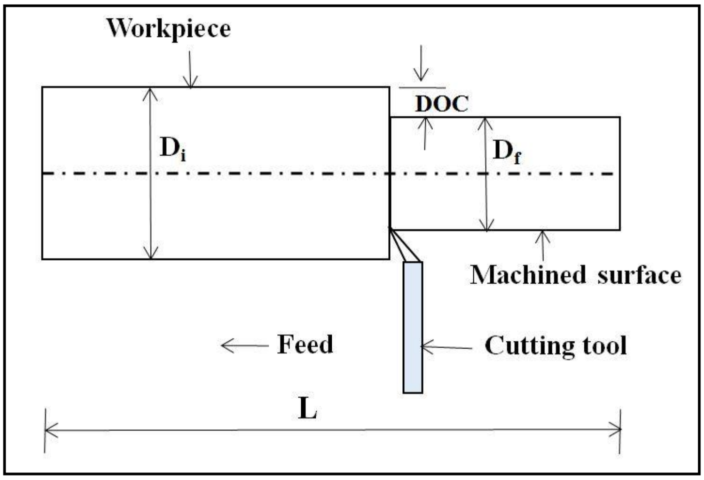


Figure 1.3 Basic turning operation

CNC turning

In order to meet the ever increasing requirements of low cost high quality products, higher production rate, enhanced dimensional accuracy, lower surface finish etc., the conventional turning operations are being continuously substituted by the computer numerical control (CNC) technology. The achievable precision and accuracy through adoption of CNC turning operations cannot be obtained by the traditional material removal processes performed on the lathe. A typical CNC turning setup, as shown in Figure 1.4, consists of a spindle attached at one end of the machine drive system while the other end accepts the chuck holding the workpiece to be turned. The most basic components in a CNC turning lathe are chuck, spindle, spindle motor, transmission (belt drive, lead screw etc.), cutting tool, tool holder and servomotor. A spindle motor powers the spindle, which helps in rotating or turning the workpiece at high speeds. The cutting tool is fitted to the tool post, driven by a feed screw transmission mechanism, like a lead screw or a ball screw. CNC turning is a typical subtractive machining process which uses a single-point cutting tool to remove material from external surface of the workpiece. However, the cutting tool is non-rotary while the workpiece rotates around it. This machining process uses computer generated programs and codes to define the position of the tool and achieve the desired shape. As the workpiece rotates, the cutting tool continues to make cut till the workpiece reaches the desired programmed shape. Modern CNC turning machines have various tools with different sizes and shapes, and spindles with varying speed capabilities, resulting in achieving wider range of geometries with higher dimensional accuracy.

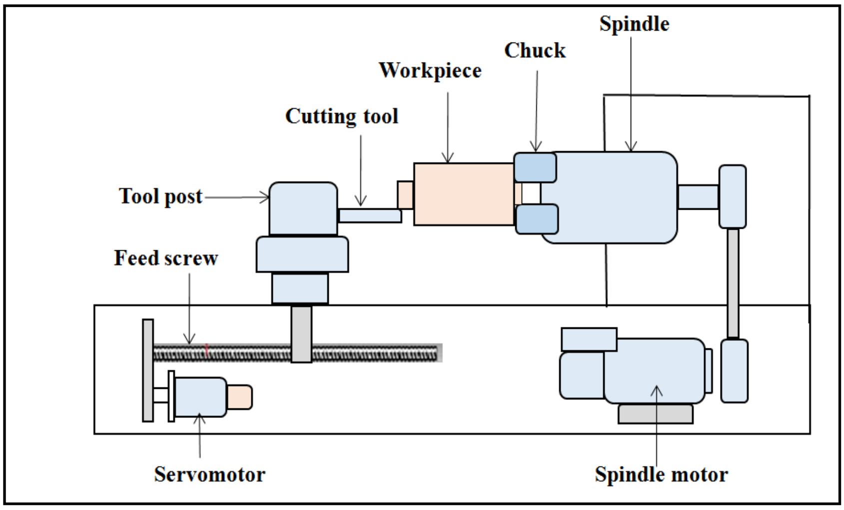


Figure 1.4 A typical CNC turning operation

EDM process

The electrical discharge machining (EDM) is a thermal NTM process in which the material is removed by melting and vaporization through a series of sparks occurring between two electrodes, i.e. anode and cathode in the form of workpiece and tool, respectively, immersed in a dielectric medium. Pulse DC power is applied across the tool and the workpiece, and a suitable voltage is built up across the electrodes. A very small gap between the electrodes is constantly maintained to establish electrostatic field which ultimately generates electric sparks during the machining operation, as depicted in Figure 1.5. The temperature of the area under spark is in the range of 8000˚C to 12000˚C or as high as 20000˚C [8], resulting in partial melting and vaporization of the work material. The debris particles accumulated in the electrode gap during machining are flushed away by the dielectric liquid which is supplied through a nozzle. EDM is popularly employed to machine electrically conductive materials of any hardness, having a distinct advantage in the production of moulds, dies and automotive components. Additionally, EDM avoids direct contact between the workpiece and the electrode, thereby eliminating mechanical stresses and vibrations during actual machining. As a result, it can economically machine various difficult-to-cut materials having complex shapes, providing significant support to the manufacturing industry.

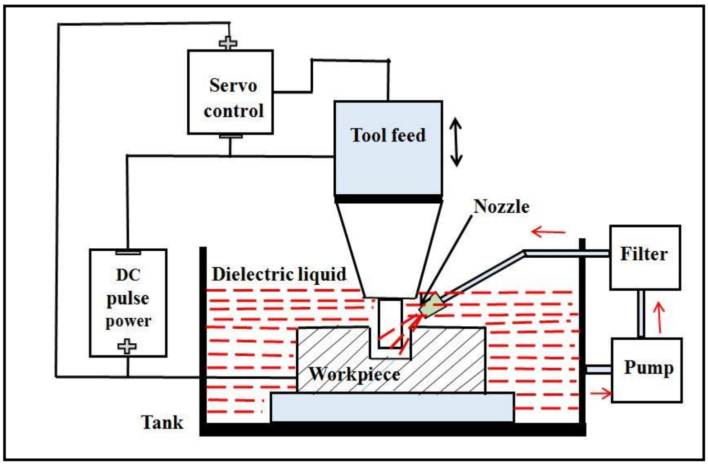


Figure 1.5 An EDM process

WEDM process

The wire electrical discharge machining (WEDM) process is one of the preferred material removal methods, capable of machining conductive HSTR materials while generating complex shapes and profiles [9]. In this process, a continuously moving thin wire, made of brass, copper or zinc having 0.002 to 0.005 mm diameter, is passed through the workpiece, thereby eliminating the need of preshaped electrodes, as required in EDM process. The movement of the wire is controlled by CNC mechanism to attain the precise geometry required for the workpiece. During WEDM operation, high current is discharged from the wire to the workpiece maintaining a small spark gap through the dielectric fluid. In this process, material is removed due to erosion effect, producing a series of rapid, repetitive spark discharges in a short span of time between the wire (tool) and the workpiece which are both adequately immersed in a dielectric medium. The debris formed during the WEDM operation is flushed away from the machining zone with a stream of dielectric liquid being supplied with the help of a nozzle in the WEDM setup, as shown in Figure 1.6.

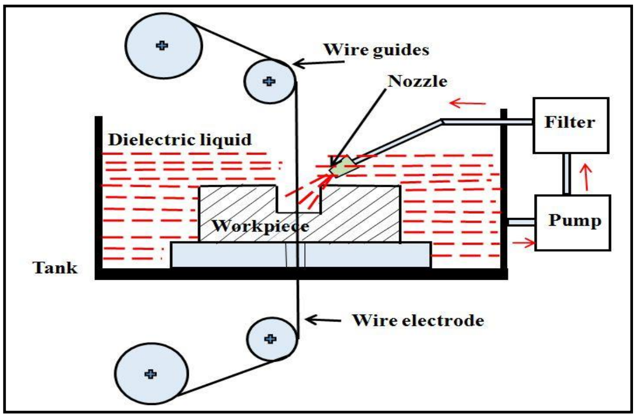


Figure 1.6 A WEDM process

ECM process

The electro-chemical machining (ECM) process is based on the principle of Faraday's law of electrolysis. In this process, workpiece is the anode and the pre-shaped tool acts as the cathode. As exhibited in Figure 1.7, a DC or pulse voltage (10-40 V) is applied across the workpiece and the tool, and an electrolyte solution flows with a very high velocity within the inter-electrode gap (0.05-0.6 mm). During this process, the cathode tool moves towards the workpiece and the material gets dissolved from the workpiece due to electrolytic action forming the desired shape on the workpiece surface. Because it is a contactless machining process, many difficult-to-cut materials can be effectively machined regardless of their hardness and strength. It is widely employed in the production of semi-conductor devices, and also in aerospace, automobile and medical equipment industries. Moreover, higher material removal rate (MRR) with excellent precision as well as good surface quality is the specific characteristic of this process, but its application is limited to electrically conductive materials only.

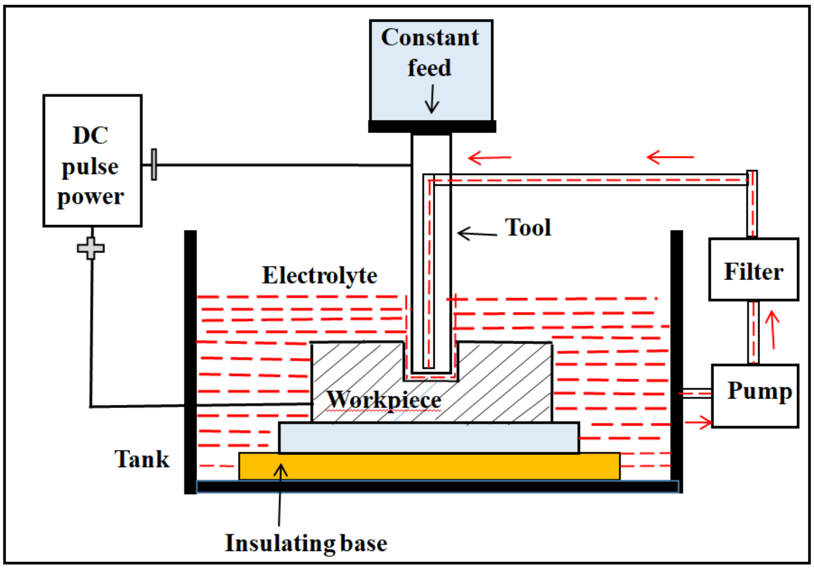


Figure 1.7 An ECM process

USM process

It is a mechanical type of NTM method used to machine conductive as well as non-conductive materials possessing high hardness above 40 HRC, such as ceramics, quartz etc. In ultrasonic machining (USM), a shaped tool, a transducer and abrasive slurry are used for material removal, as represented in Figure 1.8. In this process, material is removed from the workpiece mainly due to mechanical abrasion caused by hammering action of the abrasive particles. During USM process, the transducer converts high frequency electrical energy into linear mechanical motion (or high frequency mechanical vibration). These high frequency vibrations are then transmitted to the tool. The tool vibrates at an ultrasonic frequency of 20 kHz or above, with amplitude ranging from 25 to 75 µm. Water mixed with abrasive particles is delivered into the machining zone. As the tool vibrates in the downward direction towards the workpiece, it strikes the numerous abrasive particles with high kinetic energy and hits the work surface with large force, sufficient to remove material from the workpiece surface.

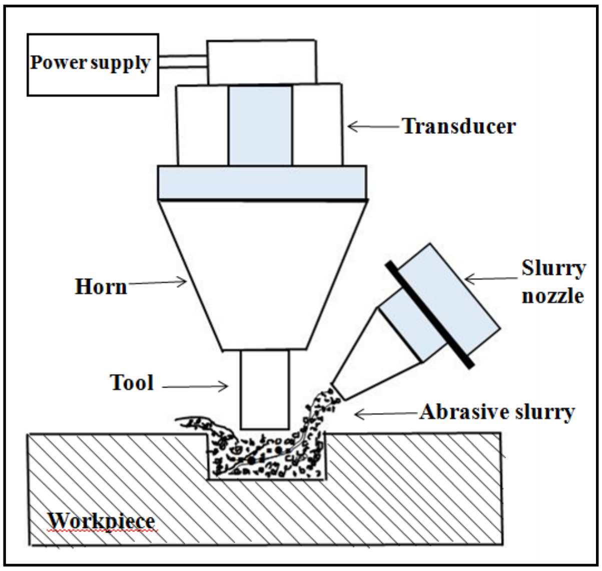


Figure 1.8 Fundamental principle of an USM process

1.3 A genesis of data mining

During the 1960's and 1970's, datasets grew in size and complexity allowing for terabytes of data to be stored. As a result, statistical technique of decision trees (DTs) was used to predict the outcomes from the massive data for the first time. Once the primitive computers had been developed in 1970's, genetic algorithms (GAs) were outlined. These algorithms were the basis of beginning of data mining (DM) that was developed in late 1980's and became the established discipline within the scope of computer science. Then by early 1990's, DM had become to be recognized as one of the processes of Knowledge Discovery of Databases (KDD). It is thereby one step in KDD which represents the entire process of turning low level data into useful knowledge [10]. DM is viewed as a natural progression of information technology, and a convergence of various related disciplines and application domains. It is widely used in marketing and sales, with popular applications, including customer relationship management and market basket analysis. It is also widely employed in genetics, bio-informatics and medicine in recent years. Hence, DM emerged during the late 1980's and advanced significantly during 1990's, and continues to flourish in a new era of data age. The application of DM in manufacturing started in the 1990's [11-13] and has since become a field of increasing interest. DM has a great potential in manufacturing, since it transforms the raw data into useful knowledge. In addition to enabling preventive maintenance and prediction of machine failure, it can provide deeper insights into various processes. Hence, DM can uphold quality while reducing costs due to damages or loss of production. There is a growing interest and a recent trend to employ DM in the manufacturing area. On the other hand, it is important to mention that DM in the field of manufacturing is a demanding task due to various factors, such as complexity of the processes, missing or incomplete data and requirement for developing accurate models with good prediction performance.

Over the decades, DM has been utilized in various but restricted manufacturing domains. However, research on DM in manufacturing has surged exponentially in recent years. Current DM approaches in manufacturing mainly address various fields of applications, like quality analysis correlating product quality with system parameters, specifically machine settings, to identify error causes responsible for deteriorating quality of the products [14, 15]; failure analysis of production resources, such as machines, to identify and analyze error causes and prevent future breakdowns [16, 17]; optimized maintenance planning [18, 19] etc. Multiple existing approaches are based on quality and failure analysis in semi-conductor industry due to its high automation levels and diverse factors affecting product quality [20]. It has been found that integration of DM tools with manufacturing systems especially in the domain of machining processes needs attention from the research community.

1.4 Need for application of data mining tools in machining

In the modern manufacturing environment, large amount of data from different sources of activities are collected and stored in the database system. This data is related to product design, machines, manufacturing processes, material selection, maintenance, machining performance etc., including trends, association and inter-dependencies resulting in more complexity for their interpretation. In these environments, volume of data increases with an unprecedented rate. Therefore, analysis of the manufacturing data under these conditions has become constrained due to overwhelming volume of multivariate data leading to the problem of 'rich data but poor information' [21]. The difficulty posed by an increasing volume of data in manufacturing calls for requirement of research on the tools to identify distinctive characteristics and relationships within the data. As a result, the field of DM has developed to offer automated tools and techniques employing sophisticated algorithms for data analysis and knowledge discovery, especially in the manufacturing domain.

Machining is a universal process which exists in every manufacturing field. It is performed to remove unwanted material from the workpiece surface to attain the desired shape of the final product/component while fulfilling the customers' needs. All these machining processes have several input (controllable) parameters. The performance of any machining operation can usually be characterized by the settings of these process parameters at different levels and the corresponding outputs (responses). It has been observed that there exist complex relationships between the process parameters (inputs) and responses (outputs) in machining operations. Understanding these relationships, and proposing techniques for parametric analysis and optimization in improving the performance of any process play pivotal roles in the present-day manufacturing environment. Over the decades, DM has been applied in various but limited areas of manufacturing, production and supply chain management. However, DM-based research in manufacturing has grown exponentially in the previous years.

Furthermore, due to advancement in automation and computer systems, data from the machining processes is becoming more accessible. Although, the performance of these processes has been improved while using traditional data analysis tools, newer methods and tools are now being used to mine vast volume of data produced by the computerized systems. It has now become the burden to the computers to quickly and exhaustively establish the relationships between various process parameters and responses through deployment of different DM tools and techniques. Conventional statistical data analysis tools are no longer the best alternative to be used [22]. DM tools can be highly beneficial for identifying intriguing and useful patterns in such intricated machining processes. With increasing automation and rapid development of the computational intelligence, knowledge has received significant attention in manufacturing, specially in the domain of machining processes, to build competitive advantages. Knowledge induction from data has now become extremely important in various TM and NTM processes to enhance productivity, understand process mechanism and improve future process performance. Computational intelligence, machine learning (ML) and advanced statistics-based DM approaches have produced new intelligent tools for automated extraction of useful information and knowledge. Availability of huge machining and manufacturing-related data in digital form has also accelerated application of DM tools to aid the process engineers in identifying the tentative settings of various process parameters to obtain the desired response values. The growth of DM, and diversity of its tools and techniques provide great opportunities to the researchers to attain the goal of predicting outcomes and discovering relationships in data. All these factors emphasize an urgent need to explore applications of advanced DM techniques for obtaining the optimal machining performance, and discover hidden patterns between the process parameters and responses for the considered machining processes. It is anticipated that DM would significantly influence current practices in machining. Some of the DM applications relevant to this research work are reviewed in the following sub-section.

1.5 Literature review

Various DM approaches have already been proposed by the past researchers for diverse manufacturing applications in the engineering field. Quinlan [23] proposed an approach of synthesized DTs used in various systems and described one of them, called ID3 in detail. The results from the studies showed the procedures to modify the methodology which would help to deal with noisy and/or incomplete information. The reported logical difficulties of the fundamental algorithm were discussed and two approaches to overcome the identified problem were compared with the help of illustrations. Koonce et al. [24] developed a software tool called DBMine to implement three common DM methodologies, i.e. Bacons algorithm, DT and DB-Learn. DBMine executed in Microsoft Visual Basic 3.0 would assist in utilizing data in Microsoft Access 2.0 and Watcom SQL databases. Job shop sequences developed by GA were presented as an example to validate efficacy of the adopted methodologies. Hence, major DM algorithms and tools were introduced to implement the mentioned methodologies assisting industrial engineers in DM applications. Meng and Butler [25] described an integrated methodology of experimental designs and neural network (NN) to solve multi-response problems for modeling and optimization of a metal inert gas welding process. The NN was trained using the optimal welding experimental data, tested and subsequently evaluated in terms of weld quality in an actual welding experiment. The relevant data that had been established from the experimental designs were highlighted in the case study in an actual workshop. Various welding combinations were included to investigate quality of the weld during actual welding operations. The derived results thus proved suitability and accuracy of the proposed methodology in weld modeling.

Sluga et al. [26] carried out multi-stage experiments to analyze the machinability database by adopting ML methodology in the form of DT. In the preparatory phase, manual construction of higher level attributes and grouping of similar learning examples had been performed leading to more consistent DT. Relationships between the material of the workpiece to be machined, cutting tool characteristics and cutting conditions were further studied. In order to predict tool characteristics, cutting geometry and cutting parameters from a set of attribute values, multiple DTs were developed during the learning process. Thus, the study provided deeper understanding of the machinability domain, and a potential for knowledge synthesis regarding cutting tool and workpiece material as a means of optimizing operation planning for numerical control (NC) machines and automated process planning. Maki et al. [27] developed an intelligent system based on DM approach for analysis of manufacturing data. This approach comprised three steps, i.e. a) feature extraction, b) combinatorial search and c) presentation. Later, it was effectively applied to conduct large scale integration fault analysis, showing how DM approach might be beneficial for the engineers to focus their attention on searching for faults.

Chaudhary et al. [28] reviewed knowledge discovery and DM applications in the wide field of manufacturing with a special focus on various types of DM functions, such as characterization and description, association, classification, prediction, clustering and evolution analysis to be performed on the data. It was observed that there had been a noticeable increase in the applications of DM in the field of manufacturing during the considered time period. The review also pointed out the advancements in DM applications as well as deficiencies existed in the area of manufacturing. It had been observed that certain areas of manufacturing, such as supply chain management, production planning and control, and integration of DM tools with manufacturing systems, would require further attention from the researchers. Koonce and Tsai [29] employed DM methodology to investigate patterns in the data generated by GA in scheduling operation. Sets of rule were developed by an attribute-oriented induction method to analyze the relationships between the sequence of operations and their corresponding attributes. The developed rules could further be successfully applied to such similar cases of job shop scheduling.

Greco et al. [30] presented the contribution of extended rough set theory (RST) in multi-criteria decision analysis (MCDA). Classical use of RST, ML, DM and knowledge discovery had been confined for multi-attribute classification problems. The MCDA had been equipped with new preference models consisting of decision rules. Consequently, application of the decision rule model and ability to manage inconsistent preferential information would allow MCDA to have access to a new research field. Dabbas and Chen [31] presented an integrated relational database method for modeling and collecting data on semi-conductor manufacturing from various database sources to transform it into useful reports. Those generated reports were employed to analyze the performance by tracking several important parameters in a Motorola’s wafer fabrication process, thereby improving the overall productivity. Baek et al. [32] introduced a novel DT learning method known as statistical batch-based decision tree learning (SBDL). The concept of batch-based learning was introduced to deal with large number of examples collected from the processes. A two-phase fitness test was carried out to measure the performance of DT and identify appropriate update points. Further, effectiveness of SBDL had been validated using a real dataset collected from a TFT-LCD manufacturing company.

Kim et al. [33] developed a mathematical model to predict the top-bead width in a robotic gas metal arc (GMA) process using NN and multiple regression methods. The relationships between various process parameters and top-bead width were defined along with prediction of the process parameters on top-bead width. Additional multi-pass butt welds were carried out using a series of robotic GMA welding, and the performance of multiple regression and NN models was verified with subsequent selection of the most suitable model. The results indicated that the proposed models could predict the top-bead width with reasonable accuracy, providing uniform weld quality. It was also observed that an NN model would show better estimations than the other empirical models. Li et al. [34] studied the relationships between various process variables and qualities of glass coating in a coating process of manufacturing. The NN, and classification and regression tree (CART), two non-parametric DM approaches, were utilized to attain the above objective. Furthermore, sensitivity analysis and CART variable ranking had been carried out for optimization of the said coating process.

Last and Kandel [35] introduced a new perception-based methodology, known as automated perception network for automated development of compact and interpretable models. It was applied to real-world datasets from semi-conductor industry to illustrate its unique capabilities. Future use of automated perceptions in DM and knowledge discovery was also discussed. Feng and Wang [36] applied two DM techniques, i.e. regression analysis and artificial neural network (ANN), to develop predictive models for a knurling process. Experiments had been carried out using fractional factorial design plan. Three statistical metrics, i.e. coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}) and root mean squared error (RMSE) were taken into consideration for selecting the best regression model. Hypothesis testing was also conducted to measure effectiveness of each model designed by the two DM techniques, and it was observed that the ANN model would outperform the regression analysis to predict the performance of the said knurling process.

Agard and Kusiak [37] proposed a new methodology while implementing DM algorithms in design of product families. It initially used a DM technique, i.e. clustering for customer segmentation. After selection of the set of customers, an analysis of the requirements for the product design had been performed using another DM technique, known as association rules. Finally, it was shown that the considered DM techniques would perform efficiently in design of the product families. Tseng et al. [38] introduced a novel approach to address quality assurance while predicting acceptance of the CNC-machined parts, rather than envisaging precise SR values. The RST was employed to generate rules related to process parameters affecting SR. The performance of the rule-composing algorithm and validation process were tested using historical data collected from a company over the years. The results demonstrated greater accuracy compared to statistical methods in predicting acceptance levels of SR and highlighted the practical viability of the RST approach for quality control.

Chen et al. [39] defined the root-cause machine sets identification problem to reduce the manufacturing cost and also to improve the manufacturing performance. To solve the problem effectively, root cause machine identifier (RMI) method was proposed employing association rule mining technique for analysis of the correlations between combinations of machines and defective products. Pham and Afify [40] presented an overview of various manufacturing applications of ML techniques and examined the applications in which they had been successfully deployed. Special focus had been given on inductive learning used currently or could be employed in the manufacturing domain. Recent trends and current developments in ML research were also presented. Jemwa and Aldrich [41] used support vector approach in which a mapping between process conditions and process trends had been formulated by developing classification-based DT. The ML concepts pertaining to support vector formulations were also discussed with the help of binary classification problem along with examples to clarify some aspects of the proposed methodology. Using data related to process operation, a basic simulated first order reaction occurring in continuous stirred tank reactor was utilized to illustrate the concept and demonstrate application of the proposed methodology. It was observed that the application of support vector classification would provide number of advantages over the original approach used in the nearest neighbour with a Euclidean distance measure system in terms of effective outlier detection, control of data patterns, its flexible properties etc. Öztürk et al. [42] applied DM method, particularly regression tree analysis to estimate lead time for make-to-order manufacturing. Training and test data were generated from the variations of a job shop simulation model. A reasonably small subset had been selected from large set of job shop attributes for performance estimation. DM with the selected attributes was subsequently compared with linear regression and three other techniques listed in the literature. The results indicated that the proposed DM approach combined with attribute selection procedure would perform efficiently than the other referred methods.

Kusiak [43] introduced an overview of the basic concepts of ML and DM. A framework utilizing decision-making constructs, such as decision tables, decision maps and atlases was proposed to organize and apply knowledge for decision-making in manufacturing and service applications. It would aim to enable a new data-driven paradigm crucial for the current manufacturing and service organizations. Furthermore, few DM applications in pharmaceutical, medical and industrial sectors had also been proposed with a vision to enhance those applications. Rokach and Maimon [44] presented a novel feature set decomposition technique to address data characteristics-related quality enhancement. A new algorithm, called breadth-oblivious-wrapper, was developed to discover the original set of features in several subsets by constructing the DT for each projection. The results obtained were later compared with other methods, such as Naïve Bayes and C4.5 which had recommended superiority of the proposed methodology in terms of accuracy and the *F*-measure. It led to the conclusion that feature set decomposition would effectively solve the classification problems in the area of quality assurance.

Harding et al. [45] investigated the applications of DM in production processes, operations, maintenance, fault detection, decision support and product quality improvement. Various aspects, like customer-relationship management, information integration and standardization had also been briefly discussed. The review was particularly focused on the relevance of DM's applicability to manufacturing sector. Feng and Kusiak [46] addressed theory-based applications of DM, text mining, web mining and image mining in engineering design, manufacturing and logistics engineering. Based on the findings, it was observed that applications of DM in engineering design, manufacturing and logistics engineering had been rare, requiring more research. Da et al. [47] presented a DM approach to determine the assembly sequences and reduce the risk of manufacturing defective products utilizing the production data. The knowledge extracted from large datasets was utilized to sequence modules and develop product families, thereby minimizing production fault costs. In the proposed approach, the emphasis had been given on knowledge extraction in an assemble-to-order environment to improve production quality by avoiding risky sequences. Wang [48] described various techniques of DM in manufacturing and their implementations on product design and manufacturing. Emphasis had been given on DM procedures adopted to discover and extract previous unknown manufacturing knowledge from voluminous databases yielding improvement of the product manufacturing processes. Two practical examples based on manufacturing maintenance and assembly quality were subsequently utilized to demonstrate the overall DM process. Further, the limitations and benefits of DM approaches in the context of manufacturing had also been discussed.

Wang et al. [49] presented a DM-based approach to derive product quality information from large volume of quality-based manufacturing data. The efficacy of the proposed approach was illustrated and validated by an example adopted from the literature. Additionally, the process of DM in manufacturing quality data and a framework of the knowledge support necessary for design decisions were also presented. Cai-Yan and Yu-Fa [50] analyzed application of DM in quality control and quality analysis within the context of quality management. Based on RST, an improved apriori algorithm had been proposed to discover quantitative relationships between various aspects influencing product quality. After analysis, it was found that the improved algorithm would overcome limitations of the traditional apriori algorithm for qualitative analysis. Most of the prior research works, as outlined in Table 1.1, have focused on applications of DM using different discrete DM algorithms.

Table 1.1 Literature review on applications of data mining in machining/manufacturing processes

Author(s)	Methodology / tool	Application area	Key findings	Remarks
Quinlan [23]	Synthesized DTs (ID3)	Knowledge systems	Proposed ID3, compared methods to overcome algorithmic issues.	Highlighted logical difficulties and improvements in DT methodology.
Koonce et al. [24]	DBMine tool (Bacon's algorithm, DT, DB-Learn)	Industrial engineering / job shop scheduling	Developed DBMine in Visual Basic, validated using GA-based job shop sequences.	Introduced practical DM tools for industrial engineers.
Meng and Butler [25]	Experimental Design + NN	MIG welding	Trained NN with optimal welding data, validated with actual welding experiments.	Proved NN suitability for weld modeling and optimization.

Table 1.1 Contd.

Sluga et al. [26]	DTs (manual preprocessing + ML)	Machinability database	Developed DTs for predicting tool/workpiece relationships.	Enhanced understanding of machinability and optimization of NC process planning.
Maki et al. [27]	Intelligent DM system (Feature extraction + Search + Presentation)	Manufacturing fault analysis	Applied to large-scale integration fault analysis, improved focus on fault detection.	Demonstrated benefits of DM for manufacturing fault analysis.
Chaudhary et al. [28]	Review of DM applications	Manufacturing systems	Reviewed DM functions (classification, clustering, prediction, etc.), noted growing use.	Identified gaps in supply chain, production planning, and integration with manufacturing systems.
Koonce and Tsai [29]	DM + Attribute-oriented induction	Job shop scheduling	Derived rules to analyze GA-generated scheduling data; applied rules successfully to similar cases.	Showed DM's role in improving scheduling efficiency.
Greco et al. [30]	Extended RST in MCDA	Multi-criteria decision analysis	Applied extended RST with decision rules to manage inconsistent preferential information.	Opened new research fields in MCDA with decision rule models.
Dabbas and Chen [31]	Integrated relational database method	Semiconductor manufacturing	Collected and transformed multi-source data into useful reports to track wafer fabrication parameters.	Improved productivity in Motorola's wafer process.
Baek et al. [32]	SBDL	TFT-LCD Manufacturing	Introduced batch-based DT learning; validated using real dataset.	Efficiently handled large datasets and improved DT updating.
Kim et al. [33]	NN + Multiple regression models	GMA welding	Predicted top-bead width, NN outperformed regression models.	Provided accurate predictions for weld quality.
Li et al. [34]	NN + CART	Glass coating process	Studied process variables affecting coating quality, performed sensitivity and CART ranking.	Optimized coating process via DM approaches.

Table 1.1 Contd.

Last and Kandel [35]	Automated perception network	Semiconductor industry	Developed compact interpretable models, tested on real datasets.	Highlighted future role of perception networks in DM.
Feng and Wang [36]	Regression analysis + ANN	Knurling process	ANN outperformed regression models for predicting knurling performance.	Validated using statistical metrics and hypothesis testing.
Agard and Kusiak [37]	Clustering + Association rules	Product family Design	Used clustering for customer segmentation and association rules for design requirements.	Efficiently applied DM for product family design.
Tseng et al. [38]	RST	CNC machined parts Quality assurance	Generated rules to predict surface roughness acceptance levels.	More accurate than statistical methods in quality control.
Chen et al. [39]	RMI + Association rule mining	Manufacturing defect analysis	Identified machine-product defect correlations.	Reduced cost and improved performance.
Pham and Afify [40]	Review of ML in manufacturing	Manufacturing systems	Surveyed ML applications with focus on inductive learning.	Presented trends and developments in ML research for manufacturing.
Jemwa and Aldrich [41]	Support vector approach + Classification-based DT	Continuous stirred tank reactor (Process modeling)	Applied SVM with DT classification for process trends; illustrated via binary classification example.	Showed advantages over nearest neighbour in outlier detection and pattern control.
Öztürk et al. [42]	Regression tree analysis	Make-to-order manufacturing	Estimated lead time using job shop simulation data; compared with regression and other techniques.	Outperformed other methods with attribute selection procedure.
Kusiak [43]	Decision tables, Decision maps, atlases (Framework)	Manufacturing and service applications	Proposed framework for organizing knowledge for decision-making.	Outlined applications in pharma, medical, and industrial sectors.
Rokach and Maimon [44]	Feature set decomposition + Breadth-oblivious-wrapper algorithm	Quality assurance / classification	Improved accuracy and F-measure compared to Naïve Bayes and C4.5.	Effective for solving classification problems.

Table 1.1 Contd.

Harding et al. [45]	Review of DM applications	Production, operations, maintenance	Reviewed DM in fault detection, decision support, and quality improvement.	Focused on DM's role in manufacturing sector.
Feng and Kusiak [46]	DM + Text mining + Web mining + Image mining	Engineering design, manufacturing, logistics	Reviewed theory-based DM applications; highlighted rarity of such research.	Called for more research in DM across engineering domains.
Da et al. [47]	DM approach for assembly sequence optimization	Assemble-to-order environment	Extracted knowledge from data to minimize production fault costs.	Improved product family development and avoided risky assembly sequences.
Wang [48]	DM techniques in manufacturing	Product design and manufacturing	Applied DM to extract knowledge for improving processes; demonstrated via maintenance and assembly examples.	Discussed both benefits and limitations of DM in manufacturing.
Wang et al. [49]	DM-based quality data analysis	Product quality information	Derived product quality insights from large datasets; proposed knowledge-support framework.	Validated approach with literature example.
Cai-Yan and Yu-Fa [50]	RST + Improved Apriori algorithm	Quality management	Discovered quantitative relationships affecting quality; improved traditional apriori.	Overcame limitations of qualitative-only analysis in quality management.

Köksal et al. [51] exhaustively reviewed the literature from 1997 to 2007 to analyze various quality tasks, such as product/process quality description, quality prediction, quality classification and parameter optimization within the context of DM applications in the manufacturing industry. The review encompassed multiple perspectives, like data handling practices, applications of DM for each quality task across different manufacturing industries, patterns used in DM methods, results obtained, and use of software in those applications. Thus, a comprehensive summary was provided on the DM practices that could be utilized in future studies. Vazan et al. [52] presented outputs of a project based on use of various DM methodologies and knowledge discoveries from the production systems. The DM models were developed utilizing the particular approaches and methodologies for defining problems based on production management system. The new knowledge was applied to the production management system and tested using simulation models of the said production system. A significant outcome of the project had been the development of a new methodology focused on DM from databases containing operational data from the production process. Jha et al. [53] developed a DM model using DT and ANN techniques to predict SR for a CNC milling process. A comparison had been performed between those two techniques to measure the prediction accuracy for SR. The ANN model could predict SR up to 93.58% accuracy, while the obtained result of prediction had been favourable with 6.42% average percentage of error. Finally, it was proved that ANN could provide better accuracy for SR prediction in the said CNC milling process. Ferreiro and Sierra [54] presented a problem based on apparition of burr which should not exceed a height of 127 μm as specified by the aeronautical guidelines and must be removed before riveting during the drilling process in the aerospace industry. Application of DM and ML techniques was employed to develop a real-time burr detection model that could be integrated into the CNC system, thereby automating and optimizing the process. Those techniques could also be utilized for other types of processes. The study also verified that although the supervised classification, ML and statistical techniques had been found to be successful and efficient to solve problems of any sector and/or source application, further research and evaluation of results in other unknown fields would be required.

Corne et al. [55] presented the inter-relationships between DM and operations research techniques. The multi-objective aspect and methods had also been discussed to emphasize emerging use of multi-objective approaches in the said context. Yacout et al. [56] introduced a new method, known as logical analysis of data, utilizing DM and pattern recognition to monitor and control the machining process of composite materials. The proposed novel approach had been used for the first time for defining machining conditions in the form of patterns to identify the acceptable products and conditions pertaining to non-acceptable products. Bastos et al. [57] proposed a predictive maintenance system in manufacturing using DM techniques, such as classification, association rule and clustering. Data collected from various industrial plants particularly related to maintenance activities was utilized for application of DM. The behaviour of the developed patterns had been identified for early accurate detection of faults in different machines. Gröger et al. [58] introduced innovative DM approaches for manufacturing process optimization, specifically indication-based and pattern-based methods, offered by the advanced manufacturing analytics platform. The effectiveness of those approaches had been demonstrated through practical case studies. Suitable DM methods as well as their implementation had also been depicted.

Bastos et al. [59] adopted rapid miner to develop prediction models from data generated during a manufacturing process and also from various maintenance activities in industrial units. The main objective had been to predict future values based on current records to assess possibility of machine failure which in return would support the maintenance teams to predict appropriate interventions for further use by the maintenance teams. Kurniadi et al. [60] implemented ANN model in remote laser welding for sheet metal assembly process. Various factors, including laser power, melting temperature, part type, thickness and welding speed had been considered as the inputs for the network to recognize the fault patterns for estimating fault type as an output. Table 1.2 reveals that most of the previous studies have attempted to employ various DM tools in various fields to explored inter-relationships of DM with operations research.

Table 1.2 A comprehensive literature review of data mining in various machining processes

Author(s)	DM/ML tool/approach	Application/Findings
Köksal et al. [51]	Review of DM applications	Exhaustive review (1997-2007) on product/process quality, prediction, classification, parameter optimization in manufacturing. Provided comprehensive summary of DM practices for future studies.
Vazan et al. [52]	DM methodologies, simulation	Developed DM models for production management systems. New methodology proposed for extracting knowledge from operational databases.
Jha et al. [53]	DT, ANN	Predicted SR in CNC milling. ANN achieved 93.58% accuracy with 6.42% error, outperforming DT.
Ferreiro and Sierra [54]	DM, ML, supervised classification	Burr detection in aerospace drilling (≤127 μm). Real-time detection model integrated into CNC system; Verified ML and statistical methods effective but need further exploration in new fields.
Corne et al. [55]	DM + operations research, multi-objective methods	Explored inter-relationships of DM with operations research. Discussed emerging role of multi-objective approaches.
Yacout et al. [56]	LAD, DM, pattern recognition	First application of LAD to machining composites. Defined machining conditions as patterns to distinguish acceptable vs. non-acceptable products.
Bastos et al. [57]	DM: classification, association rule, clustering	Predictive maintenance system for industrial plants. Identified patterns for early fault detection in machines.
Gröger et al. [58]	Indication-based and pattern-based DM methods	Advanced manufacturing analytics platform for process optimization. Effectiveness shown through case studies.
Bastos et al. [59]	RapidMiner DM models	Developed predictive models for machine failure from manufacturing and maintenance data. Supported maintenance teams with failure prediction.
Kurniadi et al. [60]	ANN	Remote laser welding in sheet metal assembly. Inputs: power, temperature, part type, thickness, speed. Output: fault type estimation.

Accorsi et al. [61] introduced a set of data analytics models and methods for decision-making, specifically in maintenance engineering to predict in advance the performance of a production system to address the associated failure. The performance of three classification models, specifically DT, random forest (RF) and NN had been compared for real-time fault prediction. The prediction accuracy of RF had been found to be slightly better than that obtained by the other methods. Ademujimi et al. [62] reviewed recent applications of fault diagnosis using various prominent ML algorithms. The areas of application included cutting tool wear in CNC machines, SR faults and wafer etching processes in semi-conductor manufacturing. In all the cases, high fault classification rates had been obtained. It was observed that most of the case studies had been limited to fault detection in a single machine, therefore diagnosis of the entire manufacturing line consisting of several machines had not been taken into consideration. Gurgenc et al. [63] proposed the application of extreme learning machine (ELM) algorithm to develop prediction model for machining time of the cycloidal gears in CNC milling machines. The developed ELM model was later compared with feed-forward and back-propagation-based ANN algorithms in terms of various statistical measures, such as RMSE, R^2 and training time. After evaluation, it was revealed that the proposed ELM model had higher prediction accuracy than the ANN-based models.

Azadi and Kolahan [64] adopted multi-objective optimization technique to specify the combination of the optimized input parameters to simultaneously achieve higher values of MRR, and minimum values SR and tool wear rate (TWR) for an EDM process. The obtained results showed higher efficiency of the proposed hybrid methodology with less than 1% error during multi-response modeling and optimization of the said EDM process. Zhu et al. [65] proposed a deep learning-based convolutional neural network (CNN) model to predict geometric deviations for different shapes and process settings in a selective laser melting process. The effective prediction of geometric deviations prior to the manufacturing process would substantially help in design optimization during additive manufacturing. Abdulateef [66] conducted 27 experiments based on full factorial design plan on a traditional turning machine using uncoated carbide tool and measured the SR values of the machined components. Furthermore, using NN back-propagation learning algorithm, a predictive model had been developed for SR. Effects of various process parameters, such as FR, cutting depth and CS on SR during the turning operation had also been investigated by conducting sensitivity analysis studies.

Khawaja et al. [67] applied response surface methodology (RSM) and an ANN model to explore the optimal process parameters, i.e. speed, FR and DOC during high-speed machining of 15CDV6 HSLA steel, while considering tool-chip interface temperature, specific energy, yield strength and percentage elongation as the responses. Furthermore, comparative analysis had been carried out between RSM and ANN results, and it was observed that ANN had exhibited more precise and accurate results in terms of higher correlation coefficient and lower RMSE than those obtained through RSM. Palanisamy et al. [68] developed NN and multiple regression models for a WEDM process to predict the grey relational grade (GRG) values. The obtained values of grey relational coefficient (GRC) from grey relational analysis (GRA) had been utilized as the input while developing the NN model to predict the corresponding multi-performance index, in terms of GRG. Comparative analysis was also carried out to measure the deviation between the actual and predicted values. The outcomes thus revealed that the NN model would offer better predictive ability over the regression models. Nguyen et al. [69] applied six ML-based models, i.e. ANN, cat boost regression (CAT), support vector regressor (SVR), gradient boosting regression (GBR), DT and extreme GBR (XGBR) to predict SR in an ultra-precision turning process. Input variables, namely, FR, DOC, spindle speed and vibration along X-, Y-and Z-axes had been considered as the most important factors influencing surface quality. Based on the evaluation error metrics, such as mean absolute error (MAE), RMSE and R^2, the predictive performance had been thoroughly assessed, indicating that XGBR and CAT models would provide greater accuracy for predicting SR values during the said ultra-turning process.

Bakhtiyari et al. [70] conducted a comprehensive review of artificial intelligence (AI) techniques, i.e. ANN, fuzzy logic (FL), metaheuristic optimization algorithms and hybrid approaches applied for modeling and optimization of laser beam machining characteristics. It was demonstrated that the AI tools could be effectively utilized for prediction and attaining the optimal parametric setting of the said process. Bhattacharya et al. [71] considered dry turning operation on a round bar of AISI 304 stainless steel to determine values of MRR, SR and cutting force (CF) based on three input parameters, i.e. CS, FR and DOC while applying four important prediction models, i.e. regression analysis, FL, ANN and adaptive neuro-fuzzy interface system (ANFIS). The relative performance of those prediction models had been contrasted with regard to some statistical metrics and the impact of the above-mentioned turning parameters on the responses had also been analyzed.

Özden et al. [72] adopted ANN technique to predict the CFs during turning operation of both unreinforced and reinforced polyamides. During the process, FR, CS, type of material and cutting tools had been considered as the input parameters. Comparative analysis had been conducted between the predicted values obtained from ANN model and experimental values with respect to R^2 and mean absolute percentage error (MAPE). It was revealed that the prediction results from the developed ANN model had good agreement with the experimental observations. Konda et al. [73] employed four ML techniques, i.e. K-nearest neighbour (KNN), DT, RF and extreme gradient boosting (XGB) to predict the fatigue crack growth rate (FCGR) of Ti64 alloy. After developing those models, comparative analysis had been carried out to evaluate the prediction performance of each of the developed models taking into account both the training and testing phases. It was concluded that XGB had provided effective predictions for FCGR in terms of least mean squared error (MSE) and higher R^2 values than the other developed models.

Elango et al. [74] applied XGBR algorithm to develop an ML model for drilling operation on CNC machine taking into consideration four important input parameters, such as spindle speed, FR, drill tool angle and hole length, and SR as the output. The obtained prediction results for SR had been further compared with polynomial model and SVR developed using the same dataset. After the critical evaluation of those three models, it was found that XGBR had the best prediction performance, followed by SVR for the said drilling operation. Wang et al. [75] demonstrated the applications of two ML techniques, i.e. XGB and RF to predict electro-sprayed particle diameter for both nano-and micro-sized particles. The modeling relationships between electro-sprayed particle diameter and the process parameters had also been studied to determine the key parameters affecting the particle size. Ganeshkumar et al. [76] applied RF, KNN and support vector machine (SVM) algorithms to predict tool wear, workpiece vibration and SR during turning of EN8 steel with TiN-coated silicon carbide tool. Three machining parameters, such as CS, DOC and FR had been considered to analyze their effects on SR and flank wear. It was noticed that RF would exhibit better prediction accuracy than the other adopted models.

Cheng et al. [77] proposed a novel hybrid kernel extreme learning machine (HKELM) algorithm with radial basis function (RBF) and arc-cosine kernel function (RBF-Arc-HKELM) to predict SR during milling operation of nickel-based superalloys, considering cutting parameters as the inputs. Comparison had been made by employing other intelligent algorithms, such as kernel extreme learning machine-based RBF, SVR, Gaussian process regression (GPR) and light gradient boosting machine in terms of MAE, and it was observed that the prediction results for SR obtained from RBF-Arc-HKELM had significant improvement by 17.82%, 15.36%, 14.16 % and 6.26 % over the other considered algorithms, respectively. Kumar et al. [78] outlined the current status of ML technique, in the context of advanced manufacturing methods, specifically additive manufacturing. Deployment of ML techniques in the present-day manufacturing industries had been discussed with main focus on design, processes and production control within the realm of additive manufacturing.

Theeda et al. [79] demonstrated the application of a single NN for optimization of multiple properties in laser powder bed fusion of SS316L. From the developed model, the optimal values of various process parameters, such as laser power, scan speed and hatch spacing had been identified to achieve the desired SR, relative density, micro-hardness and dimensional accuracy of the stainless steel components. Huang et al. [80] explored an intelligent combination of prediction system considering two materials, i.e. stainless steel and aluminium during a CNC turning operation. Back-propagation NN had been implemented to develop both single-material and multi-material prediction models in the considered CNC turning process, with SR as output and signal factors as the inputs. The prediction results demonstrated higher accuracy for multi-material model (96.74%) than single-material model

(93.75%) for stainless steel and 89.81% for aluminium. An appropriate *t*-test had also been carried out to compare the prediction errors of the developed models, which revealed that the ANN-based prediction results obtained for multi-material models had higher accuracy than the single-material model.

Özkavak et al. [81] adopted three ML models, such as CNN, ANN and RF to predict hardness and bending strength of full dense and powder metal aluminium alloys. It was observed that those models could effectively analyze changes in the mechanical properties of AA 2024 material produced through various manufacturing methods after aging at different temperatures and durations. Yang et al. [82] introduced a novel pre-process methodology called iGATE for dimensionality reduction of inputs along with a high prediction accuracy while envisaging the mechanical properties of hot-rolled steel plates using NN and XGB. It was found that XGB would possess the best prediction performance with relative error smaller than 5%. Barrionuevo et al. [83] employed various gradient boosting algorithms, like GBR, XGBR and AdaBoost to predict surface hardness and wear resistance properties for materials, commercially employed in laser-based powder bed fusion technology. Feature importance analysis was employed to analyze the contribution of the process parameters (laser power, scanning speed, layer thickness, hatch distance and material density) on the resulting micro-hardness and wear resistance of the samples produced by laser-based powder bed fusion technology.

Chakraborty et al. [84] conducted a detailed analysis based on applications of ANN in the form of predictive tools, considering three machining operations, i.e. turning, milling and drilling to extract information on types of ANN, corresponding learning algorithm and activation function, network architecture and statistical metrics. After an exhaustive review, it was revealed that ANN models had been mostly utilized by the previous researchers in turning (42.07%), milling (34.48%) and drilling operations (23.45%). Moreover, it was observed that the feed-forward neural network (FFNN) had been preferred among various ANN models, whereas, Levenberg-Marquardt (58.3%), sigmoid (31.6%) and MSE (47.2%) had been identified as the most preferred learning algorithm, activation function and statistical measure, respectively. Farooq et al. [85] employed four ML algorithms, i.e. KNN, Gaussian regression, DT and logistic regression to access their functionality while predicting power consumption (PC) during CNC machining of Inconel 718 using real experimental dataset. Four control parameters, i.e. CS, FR, DOC and flow rate, and two performance characteristics, i.e. minimum quantity lubrication and nano fluids-based minimum quantity lubrication had been considered to perform significance analysis, predictive modeling and statistical evaluation. To determine effectiveness of the developed models, comparative analysis had been conducted, which revealed that DT would provide more accurate PC prediction on the basis of five performance evaluation metrics, such as R^2, MSE, RMSE, MAE and median absolute error than the other adopted algorithms.

Du et al. [86] investigated the effects of three cutting parameters, i.e. CS, FR and DOC on output parameters, such as energy consumption, CF and SR utilizing analysis of variance (ANOVA) during machining of cast steel. Further, high-order RSM, regression tree and SVR had been applied to model the output parameters. Comparative analysis had also been carried out, and it was observed that high-order response function would be more suitable for modeling, resulting in higher precision than the other methods. Finally, a desirability function had been employed to optimize the cutting parameters of the said process. Truong et al. [87] presented a novel approach, known as Bayesian regularized ANN (BRANN) to predict tool wear during milling of Ti6Al4V. To validate accuracy of the proposed BRANN model, the obtained results had been compared with those from other five models, i.e. linear regression, SVR, multi-layer perceptron, CNN and long short term memory (LSTM) network. Additionally, impacts of various input features, training data size, hidden units, training algorithm and transfer function on the performance of the proposed model had also been investigated to check its predictive capability.

Zhao et al. [88] proposed a novel method, called ensemble learning support vector machine (ELSVM), an multi-algorithmic regression model for prediction of SR during longitudinal ultrasonic vibration-assisted grinding process. The results obtained from ELSVM model had been compared with other individual prediction models, such as improved artificial immune system, improved artificial immune particle swarm optimization (I-AISPSO), standard particle swarm optimization (SPSO) and radial basis neural network (RBNN). From the comparative analysis, it was observed that ELSVM would possess minimum average ratio of MAE while predicting SR. Kulandaiyappan et al. [89] developed general regression neural network (GRNN), SVR and relevance regression models to predict MRR and SR values for a WEDM process, while discharge current, pulse-off time, pulse-on time and servo speed had been considered as the input parameters in the model. For all the developed models, the statistical measures, like MAPE, RMSE and R^2 had been utilized to assess the prediction accuracy. Among the developed models, it was found that GRNN had excellent prediction capability as compared to other developed models. Furthermore, ANOVA had been conducted exhibiting that discharge current and servo speed would significantly affect MRR and SR, respectively.

Yau et al. [90] proposed an ANN model utilizing automatic hyper-parameter tuning and transfer learning along with architecture of inception, densenet and layer normalization to predict tool wear during a milling operation. Combination of automatic hyper-parameter tuning with a general optimizer had been utilized to attain optimal tool wear prediction. Comparison of the proposed model had been later carried out using metrics, like MAE, RMSE and R^2 to access its effectiveness. The obtained results demonstrated that the proposed model had the best prediction accuracy than the other models, such as LSTM, RF and SVR. Song et al. [91] developed a multi-target predictive model combining a novel two-step feature-integration approach and multi-kernel Gaussian process autoregressive regression (MK-GPAR) to envisage responses, such as tool wear and SR during a milling operation. The prediction results were compared with the traditional GPAR and other traditional multi-target predictive models, like gradient boosting, SVR and KNN, indicating that MK-GPAR would provide superior predictive performance across different evaluation indicators during the said milling process.

Thapaliya et al. [92] applied five ML models, e.g. DT, RF, SVR, XGBR and ANN to predict average PC and processing time during a CNC milling operation. Effects of various input parameters, such as weighted cutting length, number of axis rotations and number of travels to machine zero point in rapid traverse on average PC and processing time had also been studied. The derived results revealed that RF would be the most suitable model for the said process. Babu [93] proposed a novel approach for classifying SR utilizing a hybrid learning algorithm, i.e. CNN-XGB. Various classification models were also presented, including RF, KNN, DT and XGB. Hybrid classification models, including CNN+RF, CNN+KNN, CNN+DT and CNN+XGB had been developed, and compared to select the best SR classification model using the considered machining dataset. The results demonstrated that CNN+XGB algorithm would produce better classification accuracy than the other ensemble approaches. Table 1.3 shows applications of different DM approach in machining along with other relevant information, input parameters and responses.

Table 1.3 A comprehensive literature review of applications of data mining in machining

Author(s)	DM/ML tool/approach	Application/Findings
Accorsi et al. [61]	DT, RF, NN classification models	Compared classification models for real-time fault prediction in maintenance engineering; RF slightly better.
Ademujimi et al. [62]	Review of ML algorithms	Fault diagnosis in CNC tool wear, SR faults, wafer etching; high fault classification but limited to single machines.
Gurgenc et al. [63]	ELM, ANN	Prediction of machining time in CNC milling of gears; ELM showed higher accuracy than ANN.
Azadi and Kolahan [64]	Multi-objective optimization	Optimized input parameters for EDM to maximize MRR and minimize SR and TWR; <1% error in predictions.
Zhu et al. [65]	Deep learning CNN model	Predicted geometric deviations in selective laser melting; useful for design optimization in additive manufacturing.
Abdulateef [66]	NN with back-propagation	27 factorial experiments on turning; predicted SR considering FR, depth of cut, cutting speed; sensitivity analysis done.

Table 1.3 Contd.

Khawaja et al. [67]	RSM, ANN	High-speed machining of HSLA steel; ANN showed higher correlation coefficient and lower RMSE than RSM.
Palanisamy et al. [68]	NN, multiple regression	WEDM process; NN predicted grey relational grade better than regression models.
Nguyen et al. [69]	Six ML models (ANN, CAT, SVR, GBR, DT, XGBR)	Predicted SR in ultra-precision turning; XGBR and CAT showed best accuracy based on MAE, RMSE, R^2.
Bakhtiyari et al. [70]	Review: ANN, FL, metaheuristics, hybrid AI	AI techniques applied in laser beam machining; effective for prediction and optimal parametric settings.
Bhattacharya et al. [71]	Regression, FL, ANN, ANFIS	Dry turning of AISI 304; compared models for MRR, SR, CF prediction; ANFIS performed best.
Özden et al. [72]	ANN	Turning of polyamides; predicted CF; ANN showed good agreement with experiments (R^2, MAPE).
Konda et al. [73]	KNN, DT, RF, XGB	Predicted fatigue crack growth rate of Ti64 alloy; XGB showed least MSE, highest R^2.
Elango et al. [74]	XGBR, SVR, Polynomial regression	Drilling in CNC; inputs spindle speed, FR, drill tool angle, hole length; XGBR best for SR prediction.
Wang et al. [75]	XGB, RF	Predicted electro-sprayed particle diameter; identified key process parameters affecting size.
Ganeshkumar et al. [76]	RF, KNN, SVM	Turning EN8 steel; predicted tool wear, vibration, SR; RF most accurate.
Cheng et al. [77]	HKELM (RBF-Arc), SVR, GPR, LightGBM	Milling nickel alloys; HKELM outperformed other methods with up to 17% improvement.
Kumar et al. [78]	Review (ML in advanced manufacturing)	Reviewed ML deployment in additive manufacturing for design, process, production control.
Theeda et al. [79]	NN	Laser powder bed fusion of SS316L; optimized laser power, scan speed, hatch spacing for SR, hardness, accuracy.
Huang et al. [80]	Back-propagation NN	CNC turning, stainless steel and aluminum; multi-material prediction models more accurate than single-material.
Özkavak et al. [81]	CNN, ANN, RF	Predicted hardness & bending strength of aluminum alloys; effective analysis of property changes after aging.
Yang et al. [82]	NN, XGB with iGATE preprocessing	Predicted properties of hot-rolled steel plates; XGB had $<5\%$ relative error.
Barrionuevo et al. [83]	GBR, XGBR, AdaBoost	Predicted hardness & wear in laser-based powder bed fusion; feature importance analysis on process parameters.
Chakraborty et al. [84]	Review (ANN applications)	Turning (42%), Milling (34%), Drilling (23%); FFNN most common; LM algorithm, sigmoid activation preferred.

Table 1.3 Contd.

Farooq et al. [85]	KNN, Gaussian regression, DT, Logistic regression	CNC machining of Inconel 718; DT most accurate for PC prediction (R^2, MSE, RMSE, MAE).
Du et al. [86]	ANOVA, RSM, Regression tree, SVR	Machining cast steel; modeled energy, CF, SR; high-order RSM most precise; desirability function for optimization.
Truong et al. [87]	BRANN, Linear regression, SVR, MLP, CNN, LSTM	Milling Ti6Al4V; BRANN outperformed others in tool wear prediction.
Zhao et al. [88]	ELSVM, AISPSO, SPSO, RBNN	Ultrasonic grinding; ELSVM had lowest MAE for SR prediction.
Kulandaiyappan et al. [89]	GRNN, SVR, Relevance regression	WEDM; predicted MRR, SR; GRNN best; ANOVA showed discharge current & servo speed significant.
Yau et al. [90]	ANN with hyperparameter tuning and transfer learning	Milling; predicted tool wear; ANN outperformed LSTM, RF, SVR.
Song et al. [91]	MK-GPAR, Gradient boosting, SVR, KNN	Milling; predicted tool wear & SR; MK-GPAR superior across evaluation indicators.
Thapaliya et al. [92]	DT, RF, SVR, XGBR, ANN	CNC milling; predicted power consumption and processing time; RF most suitable.
Babu [93]	CNN+XGB, RF, KNN, DT	Hybrid classification of SR; CNN+XGB had best accuracy among ensemble methods.

1.6 Objectives and scope of the present research work

In the present-day manufacturing industries, with increasing automation and rapid development of computational intelligence, knowledge induction has received significant attention, specially in the domain of machining processes, to build competitive advantages. Knowledge induction from data has now become extremely important in any machining process to enhance productivity, understand process mechanism and improve the future process performance. DM helps to quickly and exhaustively establish the inherent relationships between various process parameters and responses through deployment of different DM tools and techniques. In this research work, an attempt is put forward to implement DM for some of the widely-used machining processes to identify their most favourable parametric settings so as to achieve the target response values. DM techniques are becoming necessary for information-related applications of manufacturing and management decision-making due to large volume of data and fast computational facilities. DM is therefore a rapidly developing field gaining attention as well as importance. One application area is the machining where it can provide significant competitive advantage in achieving better outcomes with further improvement.

From the review of the past research works, it has been found that most of the applications of DM are related to the fields of quality control, fault diagnosis, manufacturing processes and production control. Applications of DM to the manufacturing area particularly machining are still underutilized and infrequently used since modern machining systems and processes are very dynamic and complicated. Therefore, it is necessary to make investigations and implications of DM techniques in the machining sector in a systematic way. Hence, in this research work, applications of various DM tools in the machining domain are proposed. From a large number of available machining processes, some of those processes, like CNC turning, WEDM, USM and EDM are considered in this research work for parametric analysis and prediction of responses using various DM tools. It is always necessary to understand the existent relationships between the process parameters and responses using mathematical models with some assumptions. Understanding these relationships and proposing techniques for parametric analysis in improving performance of various machining processes play pivotal roles in the present-day manufacturing environment. DM offloads this task to the computers, allowing them to quickly and comprehensively identify those relationships beneficial for the machining processes.

It has already been mentioned that for having enhanced performance of the machining processes, investigation of the influences of various input parameters on the responses and determination of their optimal combinations are very important. To fulfil this objective, popular DM tools in the form of DT, association rules and NN are applied to analyze the experimental data of different machining processes, i.e. CNC turning, WEDM, USM and

EDM. From the literature, it has been observed that applications of various statistical methods and AI techniques are the suitable approaches for development of predictive models for studying the complex material removal behaviour of different machining processes. Availability of huge machining and manufacturing-related data in digital form has also accelerated the application of DM tools to aid the process engineers in identifying the tentative settings of various process parameters to obtain the desired response values. DM thus provides area of computational intelligence offering new iterative and interactive techniques for processing large volume of data to explore valuable and understandable patterns hidden in the dataset.

Keeping in view the above-cited requirements, the objectives of this present research work are set as follows:

a) to investigate the application potentiality of various DM tools and techniques, such as classification, prediction, clustering, association rules, ANN etc. with respect to classification accuracy and prediction risk in obtaining parametric and prediction performance analyses for different machining processes,
b) to compare the relative performance of different DM tools, and examine their capability with respect to accuracy while analyzing contribution of various turning parameters on the responses and identify the best machining condition for a CNC turning operation,
c) to apply DM tools to study influences of varying WEDM parameters on the responses and identify their optimal combination for the considered machining process,
d) to investigate the effects of various machining parameters on the responses during an USM operation using CART analysis,
e) to develop DM and ML-based predictive models in the form of XGB, Kriging (KRG), RBF and gene expression programming (GEP) during EDM operation, and compare their relative performance to identify the most reliable model, and
f) to employ ANN-based DM approach for development of appropriate prediction models for an EDM process to analyze four popular NN models, i.e. FFNN, CNN, recurrent neural network (RNN) and LSTM, for predicting the quality characteristics during EDM operation, and to compare their relative prediction performance.

In the present research work, an attempt is thus put forward to apply various DM tools for parametric analysis, and predictive modeling techniques for assessing the performance of the considered machining processes. To strengthen their prediction performance, comparative analysis of the adopted DM tools is also carried out to draw proper quantitative relationships between the process responses and controllable input parameters of the considered machining processes. As a future direction of this research work, combining certain DM tools into a hybrid or ensemble model can lead to superior prediction performance, which can overcome limitations of the individual solution strategies. Thus, hybridization of DM tools may lead to high quality solutions as compared to the existing and traditional techniques. To obtain more accurate results, a data library with a large set of experimental runs can be utilized for training and testing purposes during applications of the DM and other predictive techniques. The applications of other yet to be popular DM techniques, like GRNN, ELSVM and MK-GPAR, and combining variety of ML techniques, may be explored as the alternative tools for prediction modeling of the considered machining processes. Figure 1.9 depicts graphical representation of research objectives and scope.

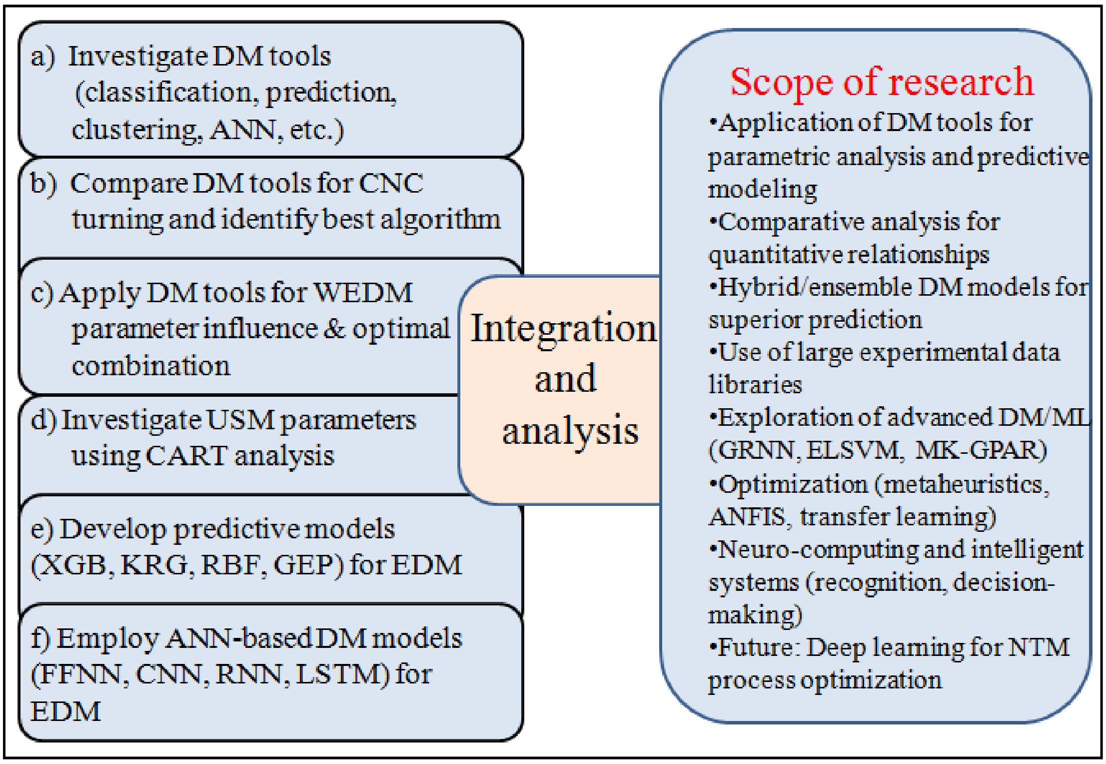


Figure 1.9 Research framework: Objectives and scope

When applying those predictive modeling techniques, it is crucial to select the most appropriate prediction tool along with its optimal architecture, tuning parameters, training algorithm and activation function, identify the best statistical metrics to evaluate its performance, and choose the optimal training data based on real-time experiments. To overcome the problems of slow convergence and over-fitting of training data, the ANN architecture may be optimized with the help of different metaheuristic algorithms. The ANFIS model may be utilized for faster data extraction. Additionally, data augmentation and transfer learning approaches can be used to deal with small volume of training data. Application of these methods may be further extended to implementation of neuro-computing principles, model optimization, and validation of outcomes using more robust statistical metrics, to develop intelligent systems that can perform diverse tasks, like image and noise recognition, autonomous decision-making and complex problem-solving. In a future study, the author would focus on some advanced applications of deep learning techniques for analyzing and obtaining the optimal parametric combinations of other NTM processes to fulfil a wide range of manufacturing requirements.

DATA MINING TOOLS

2. DATA MINING TOOLS

Since the last few decades, rapid advancements in computational facilities and information technology have enabled the human beings to move beyond manual, tiring and labourious practices for quick, straightforward and automated analysis of data. In this direction, DM has evolved out as a powerful tool with immense applicability and potentiality to turn raw data into useful information. DM occasionally known as 'knowledge discovery in databases', is the procedure of drilling through data to unveil hidden patterns/connections and anticipate future trends [94, 95]. It mainly depends on effective collection of data and warehousing as well as computer processing. The more complicated the dataset complied, the more potential it has to discover valuable insights. In engineering, databases and statistical methods are widely used because a well-analysed data forms a valuable resource that can provide significant competitive advantages and gain new insights. To identify patterns and relationships within data, statistical approaches were traditionally employed. However, due to growing complexity in machines and sensor-based equipment along with recent trends in information technology, data collecting systems and storage technology result in large amount of data, which is always increasing with wide range of attributes. The goal of DM is to solve these problems by applying mathematical models to automatically find patterns in data that has been already stored in the databases. Thus, effective development of new knowledge and information requires a close collaboration of data, domain and DM experts. The objectives and tasks of the DM process, and any factor, affecting the machining processes under study, should be identified and understood.

The application of DM has strong relations with statistics (to study data relationships), AI (to provide human-like intelligence) and ML (to learn from data to make predictions). DM thus allows to (a) filter all the chaotic and repetitive noise in the dataset, (b) predict automatic patterns based on trend and behaviour analysis, (c) envisage the likely outcomes, (d) create decision-oriented information, (e) accelerate the pace of making informed decisions, (f) focus on large dataset for analysis, and (g) form clusters based on visually documented groups of facts.

The application of DM techniques follows some procedural steps, such as (a) data cleaning (elimination of noisy and outlier data), (b) data assimilation (amalgamation of multiple data sources), (c) data selection (recovery of pertinent data from the database), (d) data conversion (transformation of data for mining purpose), (e) DM (identification of data patterns), (f) pattern assessment (extraction of attractive patterns) and (g) knowledge display (picturing of the mined information).

For the complete DM process, it is essential to clearly define the problem, specify the data scope, attributes and structure, incorporate domain knowledge, and select appropriate data and DM tools based on the target objectives. DM provides tools for analyzing large databases and uncovering patterns, trends and knowledge. Effective DM requires advanced modeling techniques combined with user-friendly interfaces to help identify meaningful and valuable relationships within the data.

2.1 Classification of data mining tools

DM encompasses techniques based on statistical analysis and AI. The models developed using these techniques can assist with issues related to detection, classification and prediction. To solve these problems efficiently, DM has emerged as a discipline that offers tools for data analysis, new knowledge discovery and analysis of large databases. For the prediction of unknown values and autonomous decision-making, powerful DM tools and techniques are available. There are various types of DM tools, based on the kind of information (attributes) known and type of knowledge sought from their applications. The purpose of DM is to identify type of the results derived from DM. According to the purpose, DM tools are mainly classified into descriptive DM and predictive DM [96].

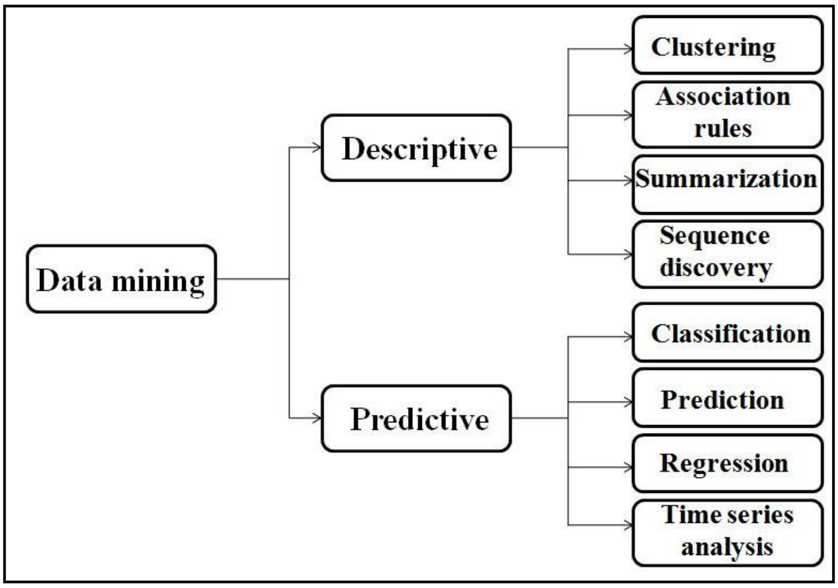


Figure 2.1 Classification of DM tools

Descriptive data mining

Descriptive DM involves exploration of understandable patterns and relationships describing a dataset. The model developed by descriptive DM provides understanding about the structure of data, describing its important features and relationships between the variables. It allows discovery of certain objects possessing similar properties and human interpretable patterns or rules from data variables. Examples of descriptive DM involve clustering, association rules, summarization and sequence discovery.

Clustering

Clustering is also known as unsupervised learning. Unlike classification (supervised learning), clustering analyzes data objects without consulting the class objects. Clustering involves classifying data into groups of similar objects where the similarity among the objects is usually measured by appropriate distance measures [95]. Clustering maps a data item into one or several clusters, i.e. natural grouping of data objects based on similarity metrics. The objects are clustered or grouped on the basis of maximizing the intra-class similarity and minimizing the inter-class similarity. In other words, data objects are grouped into clusters so that the objects within a cluster are similar to each other, but different from objects in other clusters. Generally, clustering techniques are classified in four categories, i.e.

a) Partition methods - Classify the data into *k* parts based on the distance in such a way that observations in each part are closely related to each other.

b) Hierarchical methods - Hierarchical methods group the data using bottom-up merging (agglomerative) or top-down splitting (divisive) approaches into a tree of clusters.

c) Density-based methods - They identify clusters in the data space, based on the adjacent area of high point density, separated from the other clusters by the area of low point density.

d) Grid-based methods - They involve partitioning the object space into a finite number of cells to create a grid structure, effectively organizing the objects within the grid.

Association rules

Association rule mining, introduced in 1993, is used to uncover relationships between groups of objects within a database [97]. As the name suggests, the association rules, are simple 'If-Then' statements that are employed to identify underlying relationships between independent and dependant variables in a dataset or analyze frequently occurring patterns in a dataset. These rules are applicable for non-numeric, categorical data, developed just by simple counting [98]. Since there is no class attribute in data clustering and association rule discovery, data form is unlabelled. In both the approaches, patterns are identified along one of the two dimensions of the database, i.e. attribute dimension and instance dimension. More specifically, association rules discover associations or relationships among the attributes, while data clustering determines which data instances belong together in natural groups or clusters. Summarization, also called as characterization or generalization, projects data into subsets with associated simple descriptions. It extracts information about the database, particularly by retrieving portions of data, and provides a concise and succinct summary of the data. Sequence discovery or sequential analysis is applied to determine the sequential patterns in data based on a time sequence of actions. The generated data patterns are similar to association rules but the relationship depends on time.

Predictive data mining

In the manufacturing environment, forecasting plays a key role since it can save capital by preventing severe damages when machine faults are predicted using data collected from different machines. Predictive DM uses database variables to forecast unknown or future values of other variables or attributes. This process, known as prediction, maps a data item to a real-valued prediction variable. A predictive model estimates value or characteristics of some variables based on the observed values of other variables in the database. To develop this model, a set of available data (training data) is utilized. The training set includes those variables to be predicted, called model output; and the variables used for prediction, called model input. By learning the relationships between input and output values, the model estimates a function that can predict the corresponding output value. In predictive modeling, the primary objective is to approximate the value of a specific target attribute from a set of sample training data where the attribute values are established beforehand. In [99], the main functions of various predictive DM tools are explained in detail.

Classification (also called as supervised learning) is a unique example of predictive modeling where a dataset is already segmented into pre-specified groups, and patterns are identified in the data to distinguish those groups. These explored patterns can then be adopted to categorize other datasets where the appropriate group description for the target attributes is unknown. Categorical or qualitative variables are estimated using classification models which predict the class or category of a new variable based on other known variables. The database tuples that are described by their attributes are analyzed in order to generate a model. The training dataset is made up of all the data tuples examined during the model construction. Since each training sample has the class label, this process is known as supervised learning. The learned model can be represented as classification rules, DT or mathematical formulas. Further, the model is also employed for classification of the future data or test data on the basis of classifier accuracy. DT induction, Bayesian classification, Bayesian network and NN are the examples of general DM tools used for classification. Other DM tools, such as KNN, GA, RST, FL and various hybrid techniques are also utilized for classification purpose [95].

Regression analysis is also a supervised learning technique which is applied to predict any new continuous (numeric) attribute in a dataset. It uses statistical techniques to model the relationship between one or more independent variables (whose attributes are already known) and dependent variables (which needs to be predicted). The aim is to make predictions by understanding the influence of varying independent variables on the dependent variable. Regression is an important tool in DM. Linear regression, logistic regression, polynomial regression etc. are the major types of regression method that are widely used for data analysis. Thus, the predictive DM process can have two main objectives, depending on the data and the intended outcome:

a) **Regression**: This involves predicting a numerical target attribute, aiming to estimate its value for new data.
b) **Classification**: This deals with predicting a discrete class for the outcome.

A variety of DM techniques are available to achieve these objectives, each with its own advantages and disadvantages. By considering the specific characteristics of the problem, appropriate DM tools can be chosen to extract meaningful patterns and provide valuable insights to the decision makers.

In the present research work, for parametric analysis and optimization of different machining processes, the following DM tools are considered, as presented in detail here-in-under.

2.2 Decision tree algorithm

Over the years, DT has become one of the most popular tools in knowledge discovery and DM. These tools are primarily used to explore large volume of data to uncover meaningful patterns [100]. The application of a DT algorithm aims to solve a given problem by representing it in the form of a tree structure. It basically employs development of a training model which can be able to predict class or value of target variables based on the decision rules generated from the initial dataset. It belongs to the class of supervised ML algorithm of DM which has the capability to solve both regression and classification problems. It is also a non-parametric approach having no idea about the pertaining distribution of the data. The developed DT usually follows the human thinking process with strong logic for the data to interpret. In a DT, the original dataset is broken down into smaller and smaller subsets while incrementally developing the associated DT at the same time. The final DT consists of many decision nodes and leaf nodes. A decision node may comprise two or more branches and a leaf node denotes a classification or a final decision. The top-most decision node containing the complete dataset is known as the root node. This dataset is sequentially split resulting in child nodes during classification. When no further splitting is possible, the final nodes are termed as terminal nodes. Similarly, each decision node of a DT relates to an attribute and each leaf node corresponds to a class label. In a DT, a series of 'If-Then' statements can be finally developed while tracing the path through different decision and leaf nodes, starting from the root node. The DTs have several advantages, like ability to deal with both categorical and numerical data, self-explanatory and easy to interpret, scalability with big data, capability to process datasets having errors or missing values, requirement of less computational effort, high predictive accuracy etc. The main classification methods for DT induction are CART and Chi-squared Automatic Interaction Detection (CHAID) algorithms. Both of them build DT, where each (non-terminal) node establishes a split condition to produce an optimal prediction (of continuous dependent variables) or classification (for categorical dependent variables). Thus, they can be employed to evaluate regression-type or classification-type problems. The important characteristics of a DT are outlined as below:

a) The resultant hierarchy is known as a tree and a particular section is referred to as a node.
b) The root node consists of the entire database.

c) The root node is branched conclusively forming child nodes.

d) When there is no further data classification possible, the final subgroups are known as terminal nodes or leaves.

e) There are three primary elements to be defined to implement them, i.e. a set of questions delimitating data apportionment, a criterion to institute the best division to develop child nodes and a completion rule for the classifications (stop-splitting rule).

In this research work, two DT algorithms, i.e. CART and CHAID are applied for parametric analyses of a CNC turning process and a WEDM process. In both the algorithms, each (non-terminal) node identifies a split condition which yields the optimal classification of dependent variables [101]. The details of these algorithms are presented here-in-under.

2.2.1 CART

It is a recursive partitioning method used for both regression and classification purposes. In this algorithm, the DT is constructed by splitting subsets of the data using all the predictor variables to create two child nodes repeatedly, beginning with the initial dataset. The best predictor variable is chosen based on a variety of impurity or diversity measures. The goal is to generate subsets of the data, which are as homogeneous as possible with respect to the target variable [95]. The procedural steps for this algorithm are presented as below [102]:

Step 1: The algorithm runs through the entire dataset, D to initiate the classification.

Step 2: If all the datasets in D belong to class P, generate a node P and stop, otherwise, decide a predictor F and produce a decision node.

Step 3: Split the samples in D into possible subsets using a predictor selection measure called ‘Gini index’ which is used in splitting for classification to reduce impurity of a node. Split at each node occurs only when it can generate the greatest improvement in classification accuracy.

Step 4: Identify the optimal break points while selecting the predictor in such a manner that the dependent variable serves to be the best split into two classes characterized by maximum internal uniqueness and external differentiation.

Step 5: For each split, the predictor variable with the best score of improvement is selected.

Step 6: Apply the algorithm recursively for all the subsets in D until all items or samples in a node have the same class, i.e. split is no longer possible.

Step 7: The process repeats recursively until one of the stopping rules is fulfilled:

a) If a node becomes pure, i.e. all cases in the node have identical values of dependent variable.

b) If the current tree depth reaches the user-specified maximum limit.

c) If the size of a node is less than the user-specified minimum size.

d) If the split of a node results in a child node whose node size is less than the user-specified minimum size.

Step 8: Framing of 'If-Then' decision rules, i.e. Rule: (Condition) → Y, where the condition is a combination of predictor variables and Y is the class label (decision).

In this algorithm, the measure of importance of independent variables (X) in relation to a DT is defined as the sum of improvements that X has across all the splits in the tree when it is used as primary or surrogate splitter. The importance of X is expressed in terms of a normalized quantity relative to the variable having the largest measure of importance. It ranges from 0 to 100, with the variable having the maximum importance score as 100. Thus, the variable importance plot is a good indicator which measures the importance of the independent variables (which have already appeared in the DT).

2.2.2 CHAID

This algorithm, developed by Kass [103], is a DT development approach, based on Chi-squared test, generated by repeatedly splitting the subsets into two or more child nodes starting with the initial dataset. Particularly, the predictor having the strongest relationship with the dependent variable based on *p*-value is utilized as the split node. To determine the best split at a particular node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. It is an exploratory data analysis method used to study the strongest association between a dependent variable and a large series of possible predictor variables which themselves may interact. The dependency measure may be a qualitative (nominal or ordinal) or a quantitative indicator. For qualitative (categorical) variables, a series of Chi-squared analyses is conducted between the dependent and predictor variables. For quantitative variables (continuous), *F*-test is used, where intervals (splits) are optimally determined for the predictor variables to maximize the ability to explain a dependent measure with respect to variance components [104]. This algorithm proceeds employing the following steps [103, 104]:

Step 1: The first step is to create categorical predictor variables from continuous variables by dividing the respective continuous distributions into a given number of categories.

Step 2: In the merging stage, for each dependent variable, merge non-significant categories. It determines the pair of categories that is least significant different (i.e. most similar) with respect to the dependent variables. The most similar pair is the pair whose test statistic provides the largest *p*-value with respect to the dependent variable.

Step 3: If the statistical test for the given pair of predictor categories is not statistically significant, it will merge the respective predictor categories and find the next pair of categories which may now include the previously merged categories.

Step 4: If the statistical test for the given pair of predictor categories is statistically significant, the adjusted *p*-value is computed for the merged categories by applying the Bonferroni adjustments.

Step 5: In the splitting stage, the independent or predictor variable with the lowest significant *p*-value (calculated as above) is selected as the best and the group is split on this predictor (i.e. each of the optimally merged categories of the predictor is used to define a subdivision of the parent group into a new subgroup). If no predictor has a significant *p*-value, the group is not split.

Step 6: The above-mentioned steps are repeated until all subgroups have either been analyzed or contain too few observations or cases. The stopping rules are basically the same as described in CART algorithm.

The 'If-Then' decision rules generated based on the applications of CART and CHAID classification algorithms constitute a more powerful knowledge representation to understand the effects of different input parameters on the responses for the considered machining process. When these rules are organized as a non-overlapping decision set, they become quite easy to interpret, even by a non-technical end-user. They follow a general structure, i.e. if the given machining conditions are met, then certain response values can be attained or predicted. They are probably the most interpretable prediction models, semantically resembling the natural language and human thinking process. They also provide valuable information on how to make the final decision and why certain conditions are met. The rule generation process using CART and CHAID algorithms has high speed and scalability, and is almost robust against the presence of outliers in the input dataset. The decision rules usually generate sparse models, which do not contain many features and draw final conclusions based on only a few binary statements. They can be generated from large scale datasets containing numerical and categorical information.

2.3 Association rule mining

Association rule mining is another unsupervised DM tool that aims to discover interesting correlations or inherent relationships between the independent and dependent variables in the datasets. Fundamentally, association rules, as the name proposes are 'If-Then' statements to analyze frequently occurring patterns in datasets and discover the probability of one particular event occurring as a direct result of another. The process of association rule mining is described as follows. Let,

I= {1,2,……,q}be the set of all items,

T= {1,2,……,n} be the set of instances in a database.

An association rule R is described as an expression in the form A→B, where the antecedent (A) and consequent (B) both belong to item sets, i.e. $A, B \subseteq I$. An antecedent is an item found within the dataset, whereas, consequent is an item observed in combination with the antecedent. The 'If-Then' expression thus attains a form, like 'If condition Then conclusion'. The principle used in the association rule is the conditional probability which includes two key indicators, support and confidence, for accessing the value of an association rule. The support sup(R) of the rule is the number of instances in I containing both A and B. The higher the support, the more that rule is involved in the database. The confidence of the rule R is given by:

$$\mathrm{conf(R)} = \frac{\sup(\mathrm{A \cup B})}{\sup(\mathrm{A})} \tag{2.1}$$

which is defined as the percentage of occurrences that contains item sets A which also contains item sets B. Confidence is also considered as an accuracy of the rule. Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. These thresholds can be set by the users. Additional analysis can be performed to discover interesting statistical correlations between the associated items.

Now, an association rule generation algorithm is adopted to form 'If-Then' rules from the reduced dataset with the attributes classified into appropriate number of clusters. This rule generation algorithm is presented as below [105].

Step 1: Initialize: $\mathrm{D} = \{d_1, d_2, \ldots, d_n\}$; $\mathrm{R} = \{r_1, r_2, \ldots, r_m\}$

Step 2: Evaluate $\mathrm{X}_{ij} = \mathrm{D}_i \cap \mathrm{R}_j$ for $i = 1,2,\ldots,p$; $j = 1,2,\ldots,q$

Step 3: For each $\mathrm{X}_{ij} \neq \emptyset$, a rule is generated as follows:

If $d_1 = \mathrm{V}(\mathrm{D}_i, d_1)$ and...and $d_n = \mathrm{V}(\mathrm{D}_i, d_n)$ Then $r_1 = \mathrm{V}(\mathrm{R}_j, r_1)$ and...and $r_m = \mathrm{V}(\mathrm{R}_j, r_m)$

[P, Q, C, QTY] [T]

where $\mathrm{P} = \frac{|\mathrm{X}_{ij}|}{|\mathrm{D}_i|}$; $\mathrm{Q} = \frac{|\mathrm{X}_{ij}|}{|\mathrm{R}_j|}$; $\mathrm{C} = \frac{|\mathrm{X}_{ij}|}{\mathrm{N}}$; $\mathrm{QTY} = |\mathrm{X}_{ij}|$; $\mathrm{T} = \mathrm{P} + \mathrm{Q} + \mathrm{C}$

where $\mathrm{D} = \{d_1, d_2, \ldots, d_n\}$ is the set of condition attributes, $\mathrm{R} = \{r_1, r_2, \ldots, r_m\}$ is the set of decision attributes, D_i is i^{th} equivalence class of D ($i = 1,2,\ldots,p$), R_j is j^{th} equivalence class of R ($j = 1,2,\ldots,q$), $\mathrm{V}(\mathrm{D}_i, d_k)$ are the values of condition attributes in equivalence classes of D_i, $\mathrm{V}(\mathrm{R}_i, r_l)$ are the values of decision attributes in equivalence classes of R_j, X_{ij} is the intersection of D_i and R_j, P is the percentage of objects in a current equivalence class of condition attribute set that correspond to a rule (a measure of rule confidence), Q is the percentage of objects in current equivalence class of decision attribute set that correspond to a rule, C is the percentage of objects that correspond to a rule (a measure of rule support) and QTY is the number of objects corresponding to a rule. In this algorithm, T represents the total strength (relative importance) of a rule. Higher value of T signifies more strength of a particular association rule for effective decision-making.

2.4 Predictive models

Machining is one of the most prevalent manufacturing operations that is characterized by the presence of various highly correlated input parameters. Due to their complex nature and non-linearity of metal cutting phenomena, developing an accurate analytical model using the traditional physics-based methods is highly challenging and sometimes even impossible. Nowaday’s trends are towards predictive modeling of those processes to deal with new advances in machining technologies and optimization problems with minimal computational effort and time. To achieve this, prediction performance models need to be developed and integrated with the machining processes. Development of advanced predictive models using ML and DM techniques has accelerated to accommodate these demands. A thorough evaluation of the relevant modeling techniques and their applicability for predicting complex machining operations in various industries has now become utmost important.

Predictive modeling is a statistical approach that analyzes data patterns to determine future outcomes. It is an essential aspect of predictive analytics, involving ML, AI, statistics and DM techniques to predict behaviour and trends using previously known data. The predictive modeling process involves running one or more algorithms on the datasets for which the prediction is to be carried out. It is an iterative process, and often involves training the model, employing multiple models for the same datasets, and finally selecting the best fit model. Thus, predictive modeling is a process of developing, testing and validating the model to best predict the desired output. Optimization plays an important role at the end of any predictive modeling process. Due to increasing complexity in the interrelationships among various machining performance indicators, optimization becomes essential for achieving high-quality machining performance. In this present research work, in order to forecast the desired response values of an EDM process, the following four predictive modeling techniques are considered:

a) XGB
b) RBF
c) KRG and
d) GEP

2.4.1 Extreme gradient boosting

XGB is a DT-based ML algorithm that comes under the category of ensemble learning, specifically the gradient boosting framework. DT experiences the problem of instability. To overcome this, ensemble models are utilized to provide more stability as well as predictive accuracy. Ensemble learning provides a systematic solution to combine the predictive ability of multiple learners. The aggregated outputs obtained from several models are combined into a single model as the resultant output. Bagging and boosting are the two widely used ensemble learning techniques in ML. In bagging, the DTs are paired together and the final solution is attained by considering the outputs from each of the developed DTs. In boosting, the trees are constructed sequentially with each subsequent tree aiming to minimize the errors of the previously developed tree. Each tree updates the residual errors by learning from its predecessors [106]. Boosting techniques have evolved with the development of gradient boosting models, which utilize gradient descent algorithms to more effectively and precisely reduce errors in the sequential models. XGB is based on the streamlined group calculation dependent on gradient boosting DT. It uses DTs as base learners and employs regularization techniques to improve model generalization. In addition to the usual regularization methods of gradient boosting, XGB incorporates two additional features, i.e. shrinkage and column subsampling, which help prevent data overfitting and enhance model robustness [107]. Shrinkage, similar to the learning rate in stochastic optimization, scales the newly added weights by a factor after each boosting step, reducing the influence of each tree and allowing future trees to further refine the model. Column subsampling, on the other hand, significantly decreases the computational time of the algorithm [108]. In XGB, an initial additive model is created, consisting of multiple base models. A new tree is added to this model to correct the errors produced by the previous tree, thus enhancing the model's overall performance. This iterative process continues until no further improvement is observed over a specified number of epochs, or until the number of trees in the XGB structure reaches the predefined maximum. To successfully fit the data into a XGB model, it is essential to choose the appropriate parameter values. Therefore, due to its computational efficiency, capability to analyze feature importance and effective handling of missing values, XGB is widely utilized for various tasks, including regression, classification, time series forecasting and ranking.

2.4.2 Radial basis function

RBF is a type of real valued function used in ML which depends exclusively on the Euclidean distance between the inputs and a fixed point. The fixed point can either be the origin or any other point, called centre. The 'radial' aspect refers to the circular symmetry around the centre point in the domain of the function. Thus, the value of a function at a point only depends on the radial distance from the centre, not the direction. It is a supervised ML algorithm based on RBF kernel function. The RBF kernel (also known as Gaussian kernel) is one of the most widely used kernel functions which computes a similarity score between data points based on their distance in the input space. It assigns high similarity scores to those points that are close to each other and lower scores to the points that are farther apart. The main feature of these functions is that their output decreases or increases monotonically with distance from a central point. These methods have gained significant popularity, easily surpassing the earlier RBF algorithms, due to their elegant mathematical foundations grounded in classical functional analysis, effectiveness for non-linear problems, convex nature of the optimization involved, smooth interpolation and good empirical performance [109].

2.4.3 Kriging

KRG is the geo-statistical technique used for interpolation purpose [110]. It is a method employed to estimate the value of a variable across a continuous spatial field based on a limited set of sampled data points. Unlike other basic approaches, such as inverse distance weighted interpolation, linear regression or Gaussian methods, KRG leverages the spatial correlation between the sampled points for interpolation. This approach is thus based on the spatial arrangement of the empirical observations rather than a pre-defined model of spatial distribution. Additionally, it provides estimates of the uncertainty associated with each interpolated value.

In KRG, the weights are so assigned that the points closer to the location of interest are allotted more weightage than those farther away. The method also considers point clustering, reducing the weight for clustered points to minimize bias in prediction. KRG functions as an 'optimal linear predictor' and an exact interpolator, calculating each interpolated value to obtain minimum prediction error for the considered point. For any actual sampled location, the desired value from KRG would match the observed value for that particular point and all the interpolated values would be the best linear unbiased predictors. Different types of KRG forms include ordinary KRG, simple KRG, co-KRG, universal KRG etc.

2.4.4 Gene expression programming

GEP belongs to the group of evolutionary algorithms and possesses similar features like GAs and genetic programming (GP). GEP represents linear fixed-length computer programs, which can later be converted into an expression tree. It is basically a GA, that employs populations of individuals, selects them based on their fitness, and introduces genetic diversity through application of one or more genetic operators [111]. GEP utilizes chromosomes which are composed of a linear, symbolic string of fixed length. Each chromosome contains one or more genes for encoding the expression trees. Each gene itself is a fixed-length string composed of various primitives. Likewise GP, GEP also adopts two types of primitives, i.e. functions and terminals. A function is a primitive that takes one or more arguments and produces a result after evaluation. In contrast, a terminal represents either a constant or a variable within the program. Based on the set of fitness functions, the chromosomes with the right solutions are selected and re-evaluated by genetic modifications, such as mutation, transportation and recombination. The process is repeated until the most suitable solution is attained with the required level of accuracy.

2.5 Artificial neural network

ANN is designed and developed from the biological inspired human brain which consists of a large number of interconnected elements in the form of neurons. These neurons work together to process information, recognize patterns and relationships within data and make the necessary predictions. ANN comprises three main layers, e.g. input layer representing input attributes, hidden layer depicting partial result of the application of information stored at input nodes and output layer symbolizing the predicted class. Each layer contains neurons or nodes in it for network processing. These layers are interconnected such that every single neuron in one layer is connected to all the neurons in the next preceding layer. The overall number of layers varies depending on the application. A general architecture of ANN is shown in Figure 2.2. Neurons in the input layer determine the input parameters, while neurons in the output layer represent the output parameters, whereas, data is processed in the hidden layers. By appropriately selecting the number of neurons in each layer and number of hidden layers, an effective ANN model can be established [112]. This is typically achieved through trial and error method or applying suitable optimization techniques. The key parameters for training and architecture function include learning rate, momentum, batch size, number of hidden layers, optimizer and number of neurons in the hidden layer. In an ANN structure, every neuron is connected with other neurons with specific weights that control the neuron’s output. The output of the ANN can be modified by adjusting values of these synaptic weights. In each hidden layer, neurons receive inputs, process them and generate an output signal (weight) using a net function and an activation function [113].

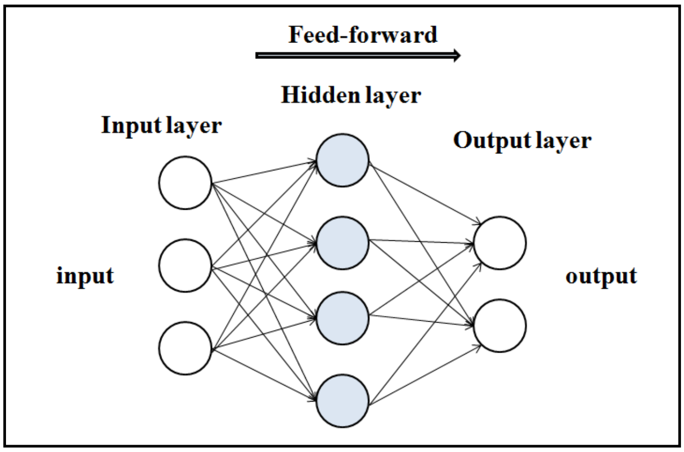


Figure 2.2 A general structure of ANN

A network collects combination of signals (input vector) to perform transformation of a nonlinear activation function using the known weights. Their output is transmitted through weighted connections to the output node. The goal of training an NN model is to define an optimal set of weights to be assigned to the neurons of the network. During training, the following steps are repeated until the network attains the predetermined accuracy or maximum number of iterations has been achieved by the model [114].

a) For each neuron, the result of applying weights and activation function to the input is computed.
b) These computations are further propagated to the output layer through the hidden layers.
c) The predicted or calculated output is compared with the expected value and the error is computed.
d) The error is back propagated through the hidden layers and the weights are recalculated.

Therefore, developing ANNs inherently begins with data cleaning and processing, followed by determining the structure of hidden layers (nodes), selecting the type of activation function, and setting learning rate of the algorithm.

2.5.1 Feed forward neural network

FFNN contains interconnected layers of neurons, where the input signal travels along the forward direction only, i.e. from the input layer towards the output layer [115]. The building block of NN is ‘neurons’ perceived as a processing unit where they are connected with each other through synaptic weight. Each neuron in a network receives ‘weighted’ information through synaptic connections from other connected neurons and generates an output by processing the weighted sum of these input signals, which come either from the environment or from outputs of other neurons [116]. A summation and an activation functions work together to produce the output. The net input from the processing neurons is computed by the summation function. On the other hand, the activation function converts the neuron’s weighted input into its output activation. It contains both linear and nonlinear algebraic equations which enable an NN to generate nonlinear relationships between the input and the output. Figure 2.3 depicts a network with a single neuron in which the input vector (multiple neurons) (i_j) is transmitted using a link that multiplies its strength by a weight (w) to make the product (w_i). This neuron also has a bias (b_j). The resulting output can be fed into other inter-connected neurons or sent directly outside the network.

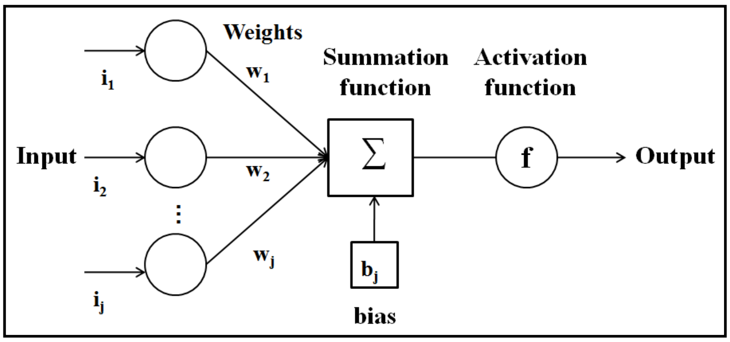


Figure 2.3 Principle of the FFNN [116]

2.5.2 Convolutional neural network

CNN is one of the most widely used deep networks with reduced connectivity. It is a deep learning model, which is designed for processing grid-patterned data, such as images, which is inspired by the structure of animal visual cortex [117] and adaptively learns the spatial hierarchies of features automatically, from low to high level patterns. The CNNs are commonly used for visual image analysis by processing data with grid-like topology. Based on ANN, CNNs transform low-level patterns into more abstract and complex representations, making them suitable for tasks, like image classification, object detection and segmentation [118]. Figure 2.4 depicts the architecture of a 1D CNN.

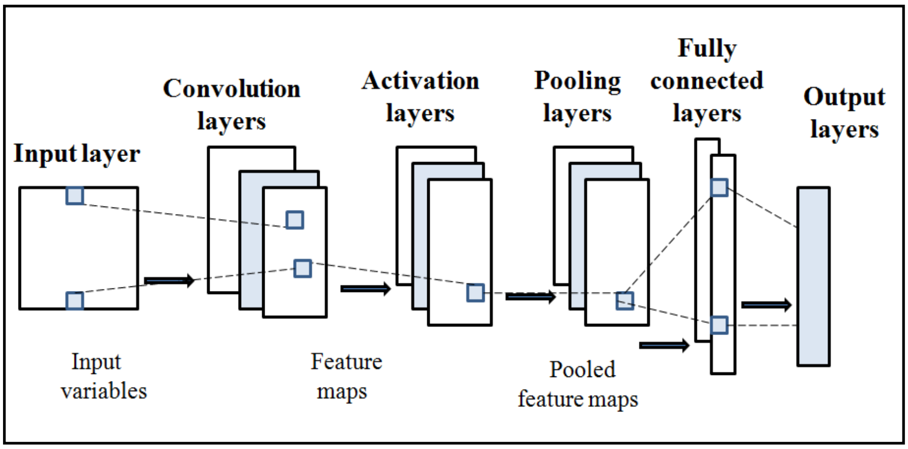


Figure 2.4 Structure of 1D CNN

The principle of CNN is based on five major types of layer, i.e. convolution layer, batch normalization layer, activation layer, pooling layer and fully connected layer.

a) Convolutional layer

A convolution layer is a fundamental element of CNN structure that performs valuable feature extraction. It combines linear and nonlinear mathematical functions, including convolution operation and activation function. This layer is composed of a series of filter kernels that performs convolution operations to generate new valuable features from the input maps. This layer is responsible for most of the heavy computational works within the network.

b) Batch normalization layer

Batch normalization layer is designed to reduce the internal covariance and speed up the training process of the deep NN. It normalizes the output of the previous layers, allowing each layer in the network to learn more independently. Batch normalization layer is typically positioned after the convolutional layer and before the activation layer.

c) Activation layer

After extraction of the feature maps, the next step is to move them to the activation layer. At this stage, nonlinear mapping function named rectified linear unit (ReLU) is utilized to enhance the CNN's representation capabilities and is usually placed after the convolution layer. Compared to other nonlinear functions, like tanh and sigmoid, ReLU can accelerate convergence of CNN more effectively and efficiently.

d) Pooling layer

The pooling layer processes each feature map independently to produce a new set of pooled feature maps with the similar number. It performs a down-sampling operation to reduce dimensions of the features and network parameters, thereby lowering computational costs and preventing overfitting. Additionally, the pooling layer compresses the input feature map, reducing its size while retaining essential information. This results in decrease in the number of learned parameters. The pooling layer, like the convolutional layer, also involves hyper-parameters, such as filter size, stride and padding.

e) Fully connected layer

The output feature maps from the pooling layer are usually flattened into a one-dimensional (1D) array or vector and fed into one or more fully connected layers, also called dense layers. In these layers, every input is connected to every output via learnable weights. This fully connected layer processes the output from the previous pooling or ReLU layer to produce an N-dimensional vector for classification. The final fully connected layer usually contains a number of output nodes equal to the number of classes [117].

2.5.3 Recurrent neural network

A RNN is a deep learning algorithm that operates in a sequential manner. While applying an NN, it is assumed that each input and output depends on all other layers. RNNs are termed 'recurrent' because their computations are carried out sequentially. Figure 2.5 represents simple architecture of a RNN.

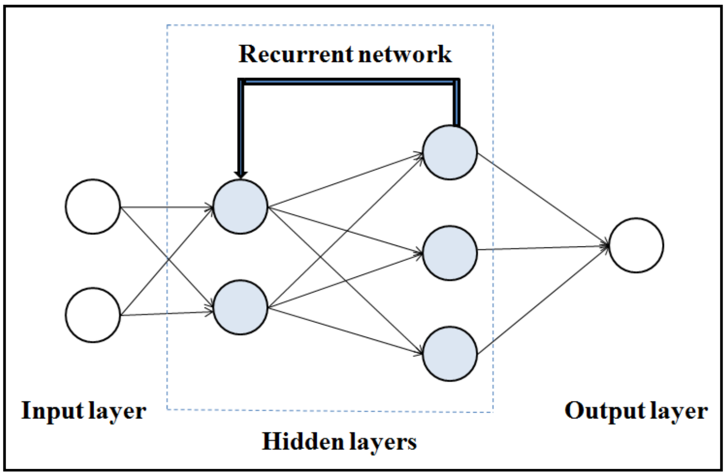


Figure 2.5 Simple structure of a RNN

In the initial stage, the recurrent network transforms independent activations into dependent ones. It also applies the same weight and bias across all the layers, simplifying the RNN and creating a standardized framework for memory. This approach allows outputs from the neurons to be fed back as inputs to the neurons in the previous layer. These layers holding the similar weights and bias, combine into a single recurrent unit. In other words, the current output is considered as an input for the next output allowing to generate more better results. Despite being more advanced than FFNNs, RNNs struggle with long input-output sequences. They are less effective at retaining information over extended periods and tend to store only minimal, less significant details from the training process. To address this issue, LSTM network has been developed as a specialized application of RNNs to tackle complex structured network problems [119,120].

2.5.4 Long short term memory neural network

LSTM is an improved version of RNN, which can effectively remember values for long or short time periods and possess the ability to forget redundant information during the analysis, which in turn, reduces occupancy of large memory. With the introduction of special gate units and a cell memory in the traditional RNNs, LSTM effectively records complex and artificial long time lag information. It consists of interacting elements, i.e. memory cell which has the ability to store the temporary state of the network, and three gates, e.g. input gate, a forget gate and an output gate that control the flow of information using the activation function to update the cell. The input gate regulates the flow of new input into the memory, while the output gate controls the flow of information from the memory to the output node. On the other hand, the flow of information from the previous memory cell to the current memory cell is controlled by the forget gate. It enables the LSTM to selectively remember or forget information from previous time steps [121]. Figure 2.6 shows simple representation of an LSTM. For each time step (t), the hidden state, h_t is updated by x_t, current data at the time step, hidden state at previous time (h_{t-1}), input gate (i), forget gate (f), output gate (o) and internal memory cell (c). LSTM also has a cell state which is exhibited by $C_{(t-1)}$ and $C_{(t)}$ for the previous and current time steps, respectively.

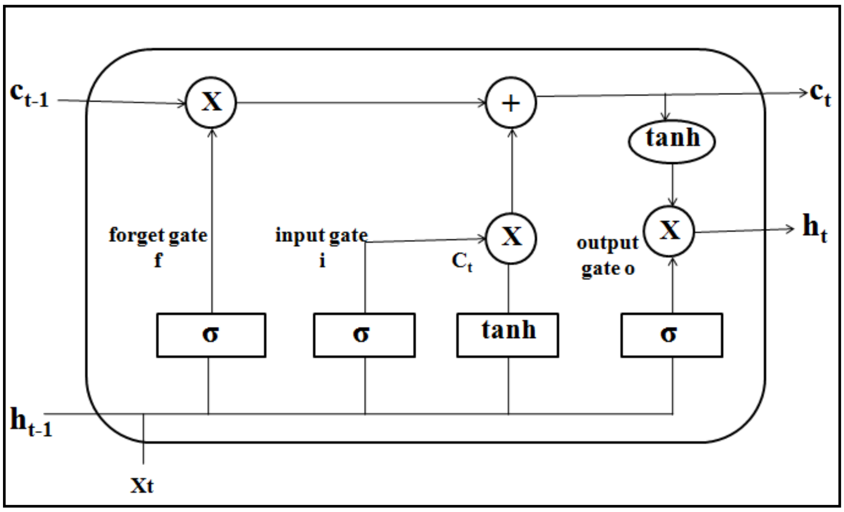


Figure 2.6 A simple LSTM representation

2.6 Statistical metrics

Statistical metrics are quantitative measures used to describe, summarize and analyze data. They provide insights into the properties and relationships within datasets, and are essential for assessing the performance of different statistical models. In this study, four such statistical metrics, i.e. MSE, RMSE, R^2 and R^2_{adj} are considered to evaluate accuracy of the adopted prediction models. The corresponding mathematical expressions of all these metrics are presented as below. Error is the difference between the predicted value and actual value of a response during machining operation, and can be considered to judge the pros and cons of a developed model. Error metrics are used to evaluate accuracy of the prediction. The MSE is a measure of the average squared difference between the predicted and actual values. It quantifies the extent to which the predictions deviate from the true values. While estimating a data sample against the actual value at a specific time, the measure of how far the data point deviates from the regression line is called the residual. However, when calculating this difference over all data points, it is known as the prediction error. Equation (2.2) provides the expression for MSE.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n} \tag{2.2}$$

where n is the number of observations, y_i represents the actual value and $\hat{y}$ denotes the predicted value of the i^{th} observation taken into consideration.

The RMSE is the root mean squared value of the deviation between the predicted and actual values. It indicates how tightly the data is clustered around the line of best fit. It is quite similar to MSE, but it describes the data more clearly. The expression for RMSE is given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n}} \tag{2.3}$$

The smaller value of RMSE indicates that the predicted output is closer to the actual output, showing higher accuracy of the model.

In regression, the most oftenly used statistical metrics to assess the degree of fit of a model are the R^2 and R^2_{adj} values, which indicate the amount of variation in a response as explained by the model. R^2 is the most important parameter used in statistical modeling to determine the prediction accuracy of a regression model which usually ranges from 0 to 1. It is a statistical measure that represents the proportion of variation in a dependent variable that can be predictable from the independent variables in a regression model. When the value of R^2 is equal to 1, it indicates that the regression model makes accurate prediction. The larger R^2 value indicates the better fit in general. The main advantage it provides is the ability to compare the developed models using different datasets. On the other hand, R^2_{adj} is a modified version of R^2 that takes into account the number of predictors in the regression model. It is noted that adding more terms to a model can increase the R^2 value. To address this issue, the adjusted R^2 is used to highlight the percentage of variation in a dependent variable that can be explained by its relationship with one or more predictor variables, while considering the number of predictors in the model. This metric represents the model's fitness and adjusts for the number of terms included. The expressions for R^2 and R^2_{adj} are provided as below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{2.4}$$

$$R^2_{adj} = 1 - \frac{\left(1 - R^2\right)\left(n - 1\right)}{n - p - 1} \tag{2.5}$$

where $\bar{y}$ is the mean of all the observations under consideration and p is the number of independent variables.

PARAMETRIC ANALYSIS OF MACHINING PROCESSES USING DECISION TREES

3. PARAMETRIC ANALYSIS OF MACHINING PROCESSES USING DECISION TREES

The performance of any of the machining operations can usually be characterized by the combination of its various input parameters and outputs (responses). These input parameters need to be more effectively controlled to achieve the target response values, minimize machining time and cost, reduce tool wear, impart more flexibility, generate complicated shapes, provide higher repeatability, attain higher dimensional tolerance, reduce power consumption etc. It has been observed that there exist complex relationships between the input parameters and responses in machining operations. Understanding these relationships and proposing techniques for parametric analysis in improving performance of the machining processes play pivotal roles in the present-day manufacturing environment.

In this section, application of a DM tool in the form of development of DT is explored to determine the optimal parametric settings of different input parameters of three machining processes, i.e. CNC turning, WEDM and USM. Two DT generation algorithms, i.e. CART and CHAID are employed here to investigate the effects of the considered machining parameters on the responses. The relative performance of both the algorithms is also compared with respect to solution accuracy and prediction risk. In this research work, the experimental data of three machining processes, i.e. CNC turning, WEDM and USM are considered for their parametric analysis employing the above-mentioned DM algorithm.

3.1 CNC turning process

CNC is a manufacturing concept where machine tools are automated to perform some predefined functions based on the instructions fed to them. CNC turning processes have found wide-ranging applications in many of the modern-day manufacturing industries due to their capabilities to produce low cost high quality parts/components with very close dimensional tolerances. In order to exploit the fullest potential of a CNC turning process, its different input parameters should always be set at their optimal levels of operation. Here, two classification tree algorithms, i.e. CART and CHAID are applied to study the effects of various turning parameters on the responses, and identify the best machining condition for the said CNC process.

Following Taguchi’s orthogonal array, Gupta et al. [122] conducted 27 experiments, focusing on five controllable parameters, such as CS, FR, DOC, tool nose radius (NR) and machining environment (E). The responses measured were TL in minutes, PC in Watts, SR in micrometers and CF in Newtons. It is worthwhile to mention here that among the considered responses, TL is the only ‘larger-the-better’ (LTB) type of quality characteristic, while, the remaining three are of ‘smaller-the-better’ (STB) type. During experimentation, each of the CNC turning parameters was set at three different operating levels, as shown in Table 3.1.

Table 3.1 CNC turning parameters and their levels [122]

CNC parameter	Symbol	Level			Unit
		Low	Medium	High	
Cutting speed	CS	120	160	200	m/min
Depth of cut	DOC	0.20	0.35	0.50	mm
Feed rate	FR	0.10	0.12	0.14	mm/rev
Nose radius	NR	0.40	0.80	1.20	mm
Environment	E	Dry	Wet	Cryogenic	-

Table 3.2 Experimental data for the CNC turning operation [122]

Exp. No.	CNC parameter					Response			
	CS	FR	DOC	NR	E	TL	PC	SR	CF
1	120	0.10	0.20	0.40	DRY	29.00	1066	1.41	171.30
2	120	0.10	0.35	0.80	WET	34.00	1560	0.71	147.50
3	120	0.10	0.50	1.20	CRYO	54.67	866	0.60	111.74
4	120	0.12	0.20	0.80	WET	34.67	1493	0.47	120.30
5	120	0.12	0.35	1.20	CRYO	51.66	987	0.19	180.60
6	120	0.12	0.50	0.40	DRY	27.00	1187	1.18	236.20
7	120	0.14	0.20	1.20	CRYO	50.00	960	0.67	157.70
8	120	0.14	0.35	0.40	DRY	24.66	1134	1.16	214.40
9	120	0.14	0.50	0.80	WET	28.33	1813	0.92	286.90
10	160	0.10	0.20	1.20	WET	27.66	1586	0.18	116.37
11	160	0.10	0.35	0.40	CRYO	47.66	1013	0.45	133.33
12	160	0.10	0.50	0.80	DRY	21.66	1240	0.43	191.23
13	160	0.12	0.20	0.40	CRYO	45.66	893	0.58	125.40
14	160	0.12	0.35	0.80	DRY	20.33	1253	0.72	149.43
15	160	0.12	0.50	1.20	WET	25.66	1773	0.31	212.46
16	160	0.14	0.20	0.80	DRY	20.00	1107	0.66	162.93
17	160	0.14	0.35	1.20	WET	22.33	1533	0.64	190.23
18	160	0.14	0.50	0.40	CRYO	41.33	1373	0.75	177.76
19	200	0.10	0.20	0.80	CRYO	40.00	1053	0.16	106.23
20	200	0.10	0.35	1.20	DRY	15.67	1373	0.23	208.50
21	200	0.10	0.50	0.40	WET	21.67	2094	0.67	209.80
22	200	0.12	0.20	1.20	DRY	14.67	1286	0.40	200.20
23	200	0.12	0.35	0.40	WET	20.33	1866	0.50	178.80
24	200	0.12	0.50	0.80	CRYO	37.66	1613	0.18	168.70
25	200	0.14	0.20	0.40	WET	18.00	1573	0.64	162.00
26	200	0.14	0.35	0.80	CRYO	34.33	1453	0.31	162.00
27	200	0.14	0.50	1.20	DRY	16.66	1667	0.48	276.16
					Minimum	14.67	866	0.16	106.23
					Maximum	54.67	2094	1.41	286.90
					Median	27.66	1373	0.58	171.30

A high speed CNC machining centre was utilized for conducting the experiments and AISI P20 tool steel bars (having diameter of 65 mm and length 275 mm) were chosen as the work material. The results of the experimental study are provided in Table 3.2. In this table, the minimum, maximum and median values for each of the responses are also shown. For parametric analysis of the considered CNC turning process and investigation of the effects of various input parameters on the process outputs (responses), the corresponding DTs are developed using CART and CHAID algorithms in SPSS 16.0 software. For arriving at the best possible solutions, various parameters of the adopted DT algorithms are fine-tuned as follows:

For CART algorithm:

Growing method: CART; Categorical dependent variables: TL, PC, SR and CF; Categorical independent variables: CS, FR, DOC, NR and E; Validation: Cross validation; Number of sample folds: 3; Growth limit: Maximum tree depth = 5; Minimum number of cases: Parent node = 3; Child node = 2; Impurity measure: Gini; Minimum change in improvement: 0.0001.

For CHAID algorithm:

Growing method: CHAID; Categorical dependent variables: TL, PC, SR and CF; Categorical independent variables: CS, FR, DOC, NR and E; Validation: Cross validation; Number of sample folds: 3; Growth limit: Maximum tree depth = 5; Minimum number of cases: Parent node = 3; Child node = 2; Significance level for: a) Splitting node = 0.03, b) Merging categories = 0.05 and c) Chi-square statistic = Pearson; Model estimation: a) Maximum number of iterations = 100; Minimum change in expected cell frequencies = 0.001 and c) Adjust significance values using the Bonferroni method.

Figure 3.1 exhibits the DT in the form of a classification tree diagram developed using CART algorithm for TL. In this diagram, TL is represented in the root node as a dependant variable. As TL is a continuous variable, its median value (27.66 min) as calculated from the experimental dataset of Table 3.2 is adopted here for the splitting purpose. For TL, being a LTB quality characteristic, its values lower than or equal to 27.66 min are termed as 'Low', whereas, values higher than 27.66 min are designated as 'High'. From the root node of the developed DT, it can be noticed that in the initial dataset, there are 13 experimental observations with high TL and 14 observations have low TL values. The first splitting is performed while taking E as the most important predictor variable. Between the two formed child nodes, node 2 appears to be a terminal node from where no further splitting can be possible. It is also identified as a pure node with no misclassification error. From node 1, taking CS as the next important predictor variable, another classification is performed with the formation of node 4 as a terminal and pure node. Two nodes, i.e. 5 and 6 now emerge out from node 3 using NR as the predictor variable. Finally, based on FR, the last two terminal nodes are constructed.

This entire classification process along with the related characteristics is provided in Table 3.3. The percentages of correct classification at all the nodes are presented in Table 3.4. It can be observed that for TL, there is no misclassification error in the DT developed using CART algorithm. When the DT of Figure 3.1 and classification characteristics are analyzed in details, several decision rules in the form of 'If-Then' statements are generated.

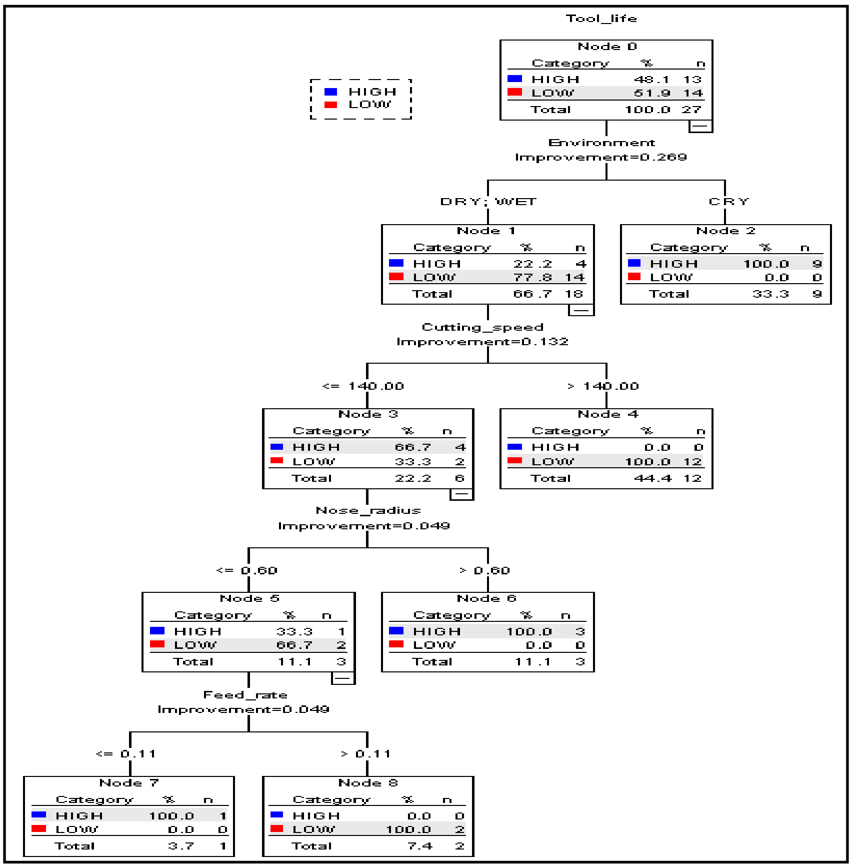


Figure 3.1 Classification tree for TL using CART algorithm

Table 3.3 Classification of TL based on CART

Classification	Node	Characteristics
First	2	Cryogenic environment provides higher TL.
Second	1,4	Dry or wet environment and CS > 140 m/min provide lower TL.
Third	1,3,6	Dry or wet environment, CS ≤ 140 m/min and NR > 0.60 mm are responsible for higher TL.
Fourth	1,3,5,7	Dry or wet environment, CS ≤ 140 m/min, NR ≤ 0.60 mm and FR ≤ 0.11 mm/rev lead to higher TL.
Fifth	1,3,5,8	Dry or wet environment, CS ≤ 140 m/min, NR ≤ 0.60 mm and FR > 0.11 mm/rev provide lower TL.

Table 3.4 Percentage of correct classification of TL based on CART

Classification	Tool life			
	Low (≤ 27.66 min)		High (> 27.66 min)	
	Number of observations	Percentage	Number of observations	Percentage
1	0	0%	9	100%
2	12	100%	0	0%
3	0	0%	3	100%
4	0	0%	1	100%
5	2	100%	0	0%

CART-based rules for TL:

Rule 1: If Environment = Cryogenic Then TL is (27.66-54.69]

[P=100%, Q=69.23%, C=33.33%, QTY=9] [T=202.56%]

Rule 2: If Environment = Dry or wet and CS > 140 m/min Then TL is [14.67-27.66]

[P=100%, Q=85.71%, C=44.44%, QTY=12] [T=230.15%]

Rule 3: If Environment = Dry or wet, CS ≤ 140 m/min and NR > 0.60 mm, Then TL is [27.66-54.69]

[P=100%, Q=23.09%, C=11.11%, QTY=3] [T=134.20%]

Rule 4: If Environment = Dry or wet, CS ≤ 140 m/min, NR ≤ 0.60 mm and FR ≤ 0.11 mm/rev Then TL is (27.66-54.69]

[P=100%, Q=7.70%, C=3.70%, QTY=1] [T=111.40%]

Rule 5: If Environment = Dry or wet, CS ≤ 140 m/min, NR ≤ 0.60 mm and FR > 0.11 mm/rev Then TL is [14.67-27.66]

[P=100%, Q=14.29%, C=7.41%, QTY=2] [T=121.70%]

where P is the percentage of objects in the condition attribute set that correspond to a rule (a measure of rule confidence), Q is the percentage of objects in the decision attribute set that correspond to a rule, C is the percentage of objects that correspond to a rule (a measure of rule support) and QTY is the number of objects satisfying a particular rule. In this algorithm, T (T = P + Q + C) represents the total strength (relative importance) of a rule [123].

Among these decision rules, rule 2, having the maximum total strength of 230.15%, states that when the said CNC turning operation is performed under dry or wet E and CS is greater than 140 m/min, the corresponding TL would be low. On the other hand, rule 1 with a total strength of 202.56 depicts that higher TL can only be achievable under cryogenic (CRYO) condition. It can also be noticed from these rules that lower FR and NR values would lead to higher TL. The importance plot for TL, as depicted in Figure 3.2, identifies E as the most important CNC turning parameter, followed by CS. NR, DOC and FR are observed to have least importance on TL.

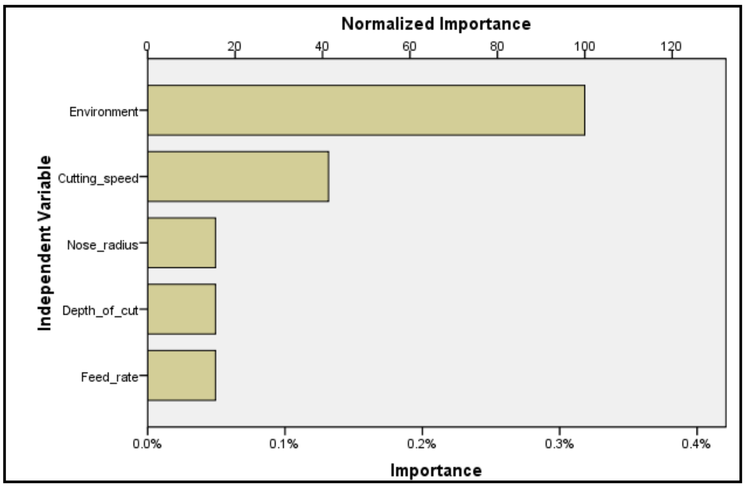


Figure 3.2 Importance plot for TL

Similarly, the related DT for TL is also generated using CHAID algorithm, as exhibited in Figure 3.3. The classification characteristics and percentage of accurate classification at the identified nodes are provided in Tables 3.5 and 3.6, respectively. As compared to five

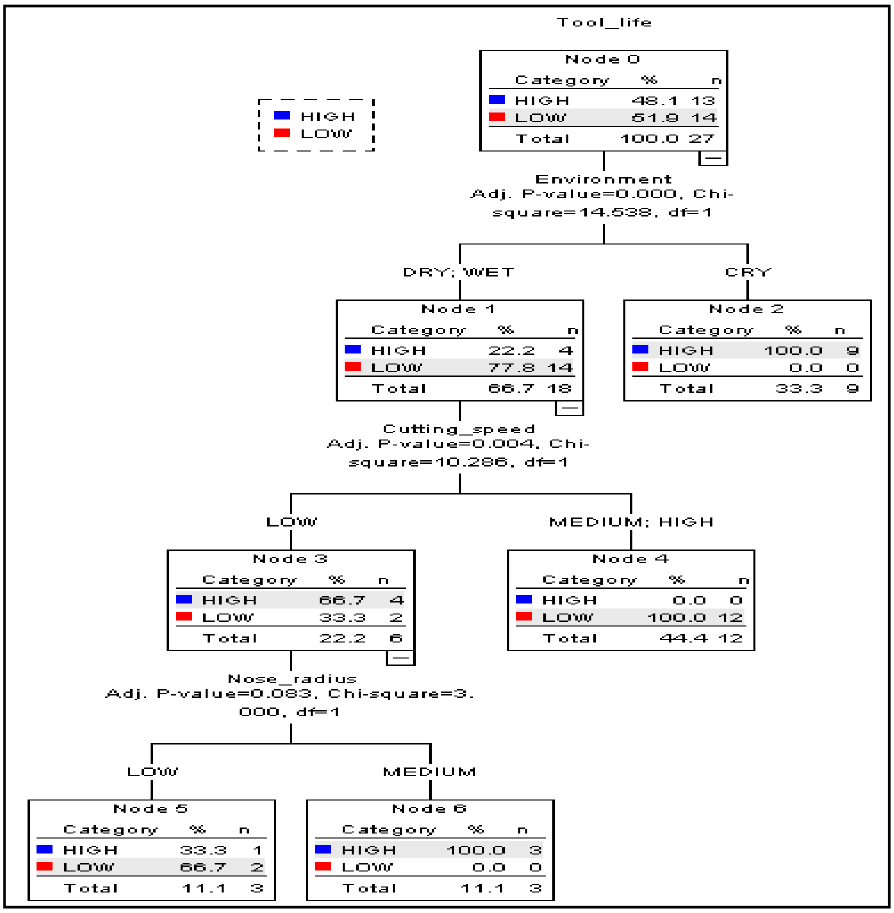


Figure 3.3 Classification tree for TL using CHAID algorithm

Table 3.5 Classification of TL based on CHAID

Classification	Node	Characteristics
First	2	Cryogenic environment provides higher TL.
Second	1,4	Dry or wet environment and medium or high CS (160 m/min or 200 m/min) lead to lower TL.
Third	1,3,5	Dry or wet environment, low CS (120 m/min) and low NR (0.40 mm) are responsible for lower TL.
Fourth	1,3,6	Dry or wet environment, low CS (120 m/min) and medium NR (0.80 mm) provide higher TL.

Table 3.6 Percentage of accurate classification of TL based on CHAID

Classification	Tool life			
	Low (≤ 27.66 min)		High (> 27.66 min)	
	Number of observations	Percentage	Number of observations	Percentage
1	0	0%	9	100%
2	12	100%	0	0%
3	2	66.7%	1	33.3%
4	0	0%	3	100%

classifications in CART algorithm for TL, there are only four classifications in CHAID algorithm. Here, at the third classification in node 5, there is 33.33% misclassification error. The corresponding rules generated using this algorithm are quite similar to those as developed by CART algorithm. Here, E and CS are observed to be the two most important CNC turning parameters affecting TL. While analyzing both these sets of decision rules, it can be concluded that for attaining higher TL, CRYO E, and low values of CS, NR and FR are always preferred. It is noteworthy to mention that DOC is found to be an insignificant parameter with no effect on TL. The ANOVA results for TL, as reported by Gupta et al. [122], also identified E as the most critical CNC turning parameter, contributing 69.27%, followed by CS with a 24.36% contribution. The DOC had almost negligible impact, contributing just 0.19% to TL. Using signal-to-noise (S/N) ratio values, the optimal parametric combination for achieving higher TL was recommended as lower CS, FR, DOC, medium NR and CRYO E which almost matches with that as proposed by the DTs.

CHAID-based rules for TL:

Rule 1: If Environment = Cryogenic Then TL is (27.66-54.69]
[P=100%, Q=69.23%, C=33.33%, QTY=9] [T=202.56%]

Rule 2: If Environment = Dry or wet and CS = medium or high Then TL is [14.67-27.66]
[P =100%, Q =85.71%, C =44.44%, QTY=12] [T=230.15%]

Rule 3: If Environment = Dry or wet, CS = low and NR = low, Then TL is (27.66-54.69]
[P=66.70%, Q=14.28%, C=7.40%, QTY=2] [T=88.38%]

Rule 4: If Environment = Dry or wet, CS = low and NR = medium, Then TL is [14.67-27.66]
[P=100%, Q=23.07%, C=11.11%, QTY=3] [T=134.18%]

The DT for PC, developed using CART algorithm, is shown in Figure 3.4. When the PC is less than or equal to 1373 W, it is designated as ‘Low’ and when it is greater than 1373 W, its value is ‘High’. As it is an STB type of response, its ‘low’ values are always preferred. The ‘If-Then’ rules extracted from the DT of Figure 3.4 highlight that dry or CRYO E and CS less than or equal to 180 m/min would always lead to lower PC (rule 2 with total strength of 224.44%). Thus, wet E is responsible for higher PC (rule 1 with total strength of 208.33%). Lower FR and DOC would cause lower PC. The importance plot of Figure 3.5 identifies E as the most critical CNC turning parameter affecting PC, followed by CS. Interestingly, NR plays no significant role on PC. The rules developed from the DT generated using CHAID algorithm also confirm these observations. Combining both the sets of decision rules from CART and CHAID algorithms, the optimal parametric mix of dry or CRYO environment, low or medium CS, and low FR and NR would always lead to lower PC. Gupta et al. [122] also observed that machining environment = CRYO, CS = low, FR = low, DOC = low and NR = medium had been responsible for attaining the most desirable value of lower PC in the said CNC turning process.

CART-based rules for PC:

Rule 1: If Environment = Wet Then PC is (1373-2094]

[P =100%, Q=75%, C=33.33%, QTY=9] [T=208.33%]

Rule 2: If Environment = Dry or cryogenic and CS ≤ 180 m/min Then PC is [866-1373]

[P=100%, Q=80%, C=44.44%, QTY=12] [T=224.44%]

Rule 3: If Environment = Dry or cryogenic, CS > 180 m/min and FR ≤ 0.11 mm/rev Then PC is [866-1373]

[P=100%, Q=13.33%, C=7.40%, QTY=2] [T=120.73%]

Rule 4: If Environment = Dry or cryogenic, CS > 180 m/min, FR > 0.11 mm/rev and DOC ≤0.28 mm Then PC is [866-1373]

[P=100%, Q=6.67%, C=3.70%, QTY=1] [T=110.37%]

Rule 5: If Environment = Dry or cryogenic, CS > 180 m/min, FR > 0.11 mm/rev and DOC > 0.28 mm then PC is [866-1373]

[P=100%, Q=25%, C=11.11%, QTY=3] [T=136.11%]

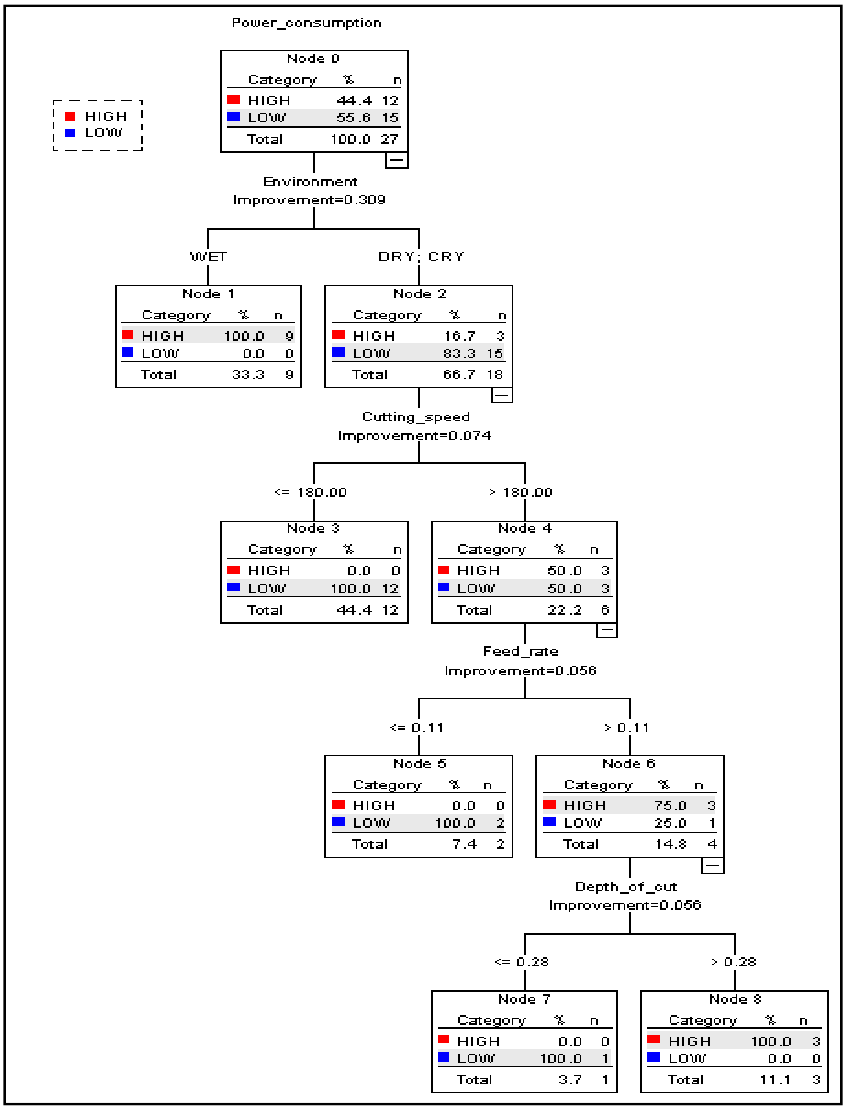


Figure 3.4 Classification tree for PC using CART algorithm

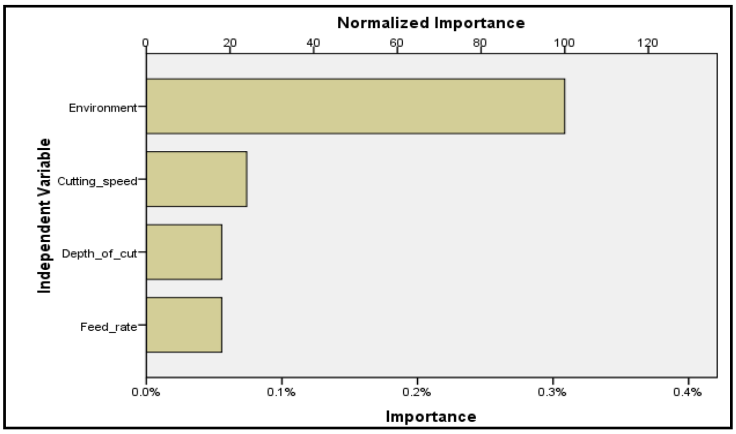


Figure 3.5 Importance of CNC turning parameters affecting PC

CHAID-based rules for PC:

Rule 1: If Environment = Wet Then PC is (1373-2094]

[P=100%, Q=75%, C=33.33%, QTY=9] [T=208.33%]

Rule 2: If Environment = Dry or cryogenic and CS = Low or medium Then PC is [866-1373]

[P=100%, Q=80%, C=44.44%, QTY=12] [T=224.44%]

Rule 3: If Environment = Dry or cryogenic, CS = High and FR = Low Then PC is [866-1373]

[P =100%, Q =13.33%, C =7.40%, QTY=2] [T=120.73%]

Rule 4: If Environment = Dry or cryogenic, CS = High and FR = Medium Then PC is [866-1373]

[P=50%, Q=6.67%, C=3.70%, QTY=1] [T=60.37%]

Rule 5: If Environment = Dry or cryogenic, CS = High and FR = High then PC is (1373-2094]

[P=100%, Q=16.67%, C=7.40%, QTY=2] [T=124.07%]

In Figure 3.6, the DT for SR developed using CART algorithm is exhibited. The corresponding 'If-Then' rules are also subsequently generated. In this case, the SR values less than or equal to 0.58 μm are termed as 'Low' (satisfactory) and those with greater than 0.58 μm values are styled as 'High'. Rule 1 with the maximum total strength of 148.40% reveals that when the machining environment is CRYO and CS is more than 140 m/min, SR of the turned components would be satisfactory (low). High NR would also provide lower SR (rule 3 with total strength of 137.52%). Similarly, high FR is responsible for poor SR. The rules extracted from the DT developed using CHAID algorithm prove that low or medium FR achieves better SR. In both the sets of rules, DOC appears as an unimportant CNC turning parameter having no impact on SR. The importance of each of the turning parameters on SR is depicted in Figure 3.7 which clearly reveals the fact that machining environment and CS are the two most significant parameters influencing SR. DOC is the least important turning parameter. Gupta et al. [122] identified that a parametric combination of CS = high, FR =

medium, DOC = medium, NR = high and machining environment = CRYO would provide better SR values of the turned components.

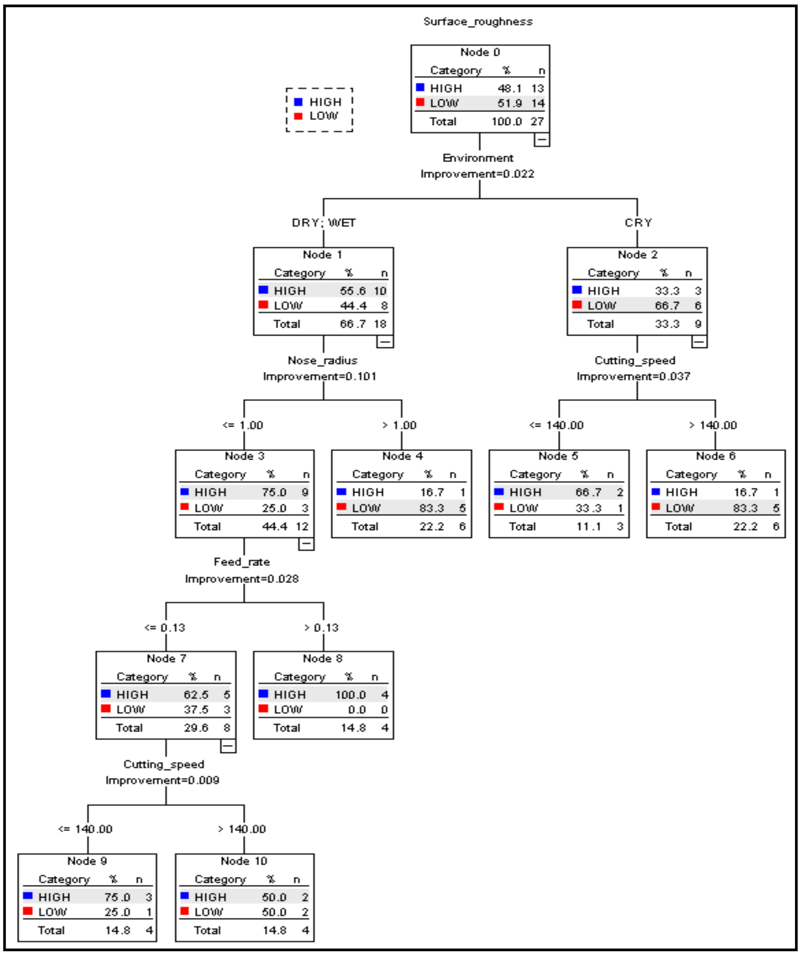


Figure 3.6 Classification tree for SR using CART algorithm

CART-based rules for SR:

Rule 1: If Environment = Cryogenic and CS > 140 m/min Then SR is [0.16-0.58]

[P=83.33%, Q=42.85%, C=22.22%, QTY=6] [T=148.40%]

Rule 2: If Environment = Cryogenic and CS ≤ 140 m/min Then SR is (0.58-1.41]

[P=66.70%, Q=15.38%, C=7.40%, QTY=2] [T=89.48%]

Rule 3: If Environment = Dry or wet and NR > 1.00 mm Then SR is [0.16-0.58]

[P=83.30%, Q=35.71%, C=18.51%, QTY=5] [T=137.52%]

Rule 4: If Environment = Dry or wet, NR ≤ 1.00 mm and FR > 0.13 mm/rev Then SR is (0.58-1.41]

[P=100%, Q=30.77%, C=14.81%, QTY=4] [T=145.58%]

Rule 5: If Environment = Dry or wet, NR ≤ 1.00 mm, FR ≤ 0.13 mm/rev and CS ≤ 140 m/min Then SR is (0.58-1.41]

[P=75%, Q=23.07%, C=11.11%, QTY=3] [T=109.18%]

Rule 6: If Environment = Dry or wet, NR ≤ 1.00 mm, FR ≤ 0.13 mm/rev and CS > 140 m/min Then SR is (0.58-1.41]

[P=50%, Q=15.38%, C=7.41%, QTY=2] [T=72.79%]

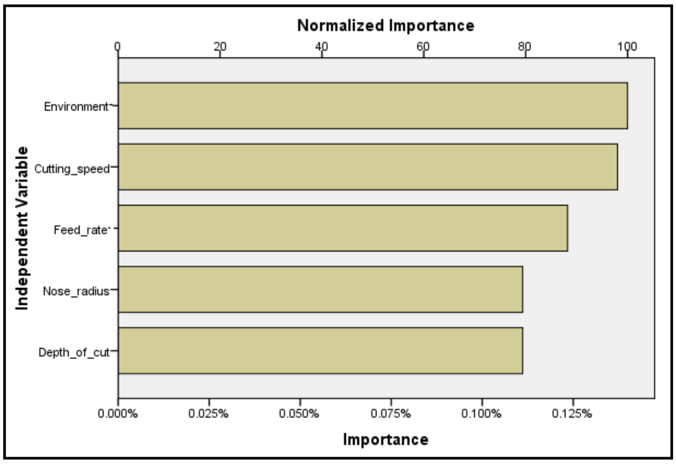


Figure 3.7 Importance plot for SR

CHAID-based rules for SR:

Rule 1: If Environment = Cryogenic and CS = High Then SR is [0.16-0.58]

[P=100%, Q=21.43%, C=11.11%, QTY=3] [T=132.54%]

Rule 2: If Environment = Cryogenic, CS = Low or medium and FR = High Then SR is (0.58-1.41]

[P =100%, Q =15.38%, C =7.41%, QTY=2] [T=122.79%]

Rule 3: If Environment = Cryogenic, CS = Low or medium and FR = Low or medium Then SR is [0.16-0.58]

[P=75%, Q=21.43%, C=11.11%, QTY=3] [T=107.54%]

Rule 4: If Environment = Dry or wet, NR = High and CS = High Then SR is [0.16-0.58]

[P=100%, Q=21.43%, C=11.11%, QTY=3] [T=132.54%]

Rule 5: If Environment = Dry or wet, NR = High and CS = Medium then SR is [0.16-0.58]

[P=66.70%, Q=14.28%, C=7.41%, QTY=2] [T=88.39%]

Rule 6: If Environment = Dry or wet, NR = Low or Medium and FR = High then SR is (0.58-1.41]

[P=100%, Q=30.77%, C=14.81%, QTY=4] [T=145.58%]

Rule 7: If Environment = Dry or wet, NR = Low or medium and FR = Low or medium then SR is (0.58-1.41]

[P=62.50%, Q=38.46%, C =18.52%, QTY=5] [T=119.48%]

The DT for CF originated from CART algorithm is shown in Figure 3.8. The corresponding 'If-Then' rules are subsequently generated from this DT. When the values of CF are less than or equal to 171.30 N, they are denoted as 'Low' and when its values are greater than 171.30 N, they are designated as 'High'. An analysis of these rules reveals that when the machining environment is cryogenic, CS is less than or equal to 180 m/min and FR is less than or equal to 0.11 mm/rev, the achievable CF would be low. Similarly, a high DOC would lead to higher CF. The rules extracted from the DT based on CHAID algorithm state that cryogenic environment would always provide lower CF. On the other hand, low DOC and low or medium NR are often responsible for attaining lower CF. When the relative importance of all the considered CNC turning parameters is plotted in Figure 3.9, it identifies DOC as the most significant one influencing CF, followed by machining environment and FR. NR appears as an insignificant CNC turning parameter for CF. An optimal parametric mix of moderate CS, low FR, low DOC, moderate NR and CRYO environment had been identified by Gupta et al. [122] for lower CF, which almost corroborates with the DTs-based observations.

CART-based rules for cutting force:

Rule 1: If Environment = Cryogenic and CS > 180 m/min Then CF is [106.23-171.30]

[P=100%, Q=21.43%, C=11.11%, QTY=3] [T=132.54%]

Rule 2: If Environment = Cryogenic, CS ≤ 180 m/min and FR ≤ 0.11 mm/rev Then CF is [106.23-171.30]

[P=100%, Q=14.28%, C=7.41%, QTY=2] [T=121.69%]

Rule 3: If Environment = Cryogenic, CS ≤ 180 m/min and FR > 0.11 mm/rev Then CF is [106.23-171.30]

[P=50%, Q=14.28%, C=7.41%, QTY=2] [T=71.69%]

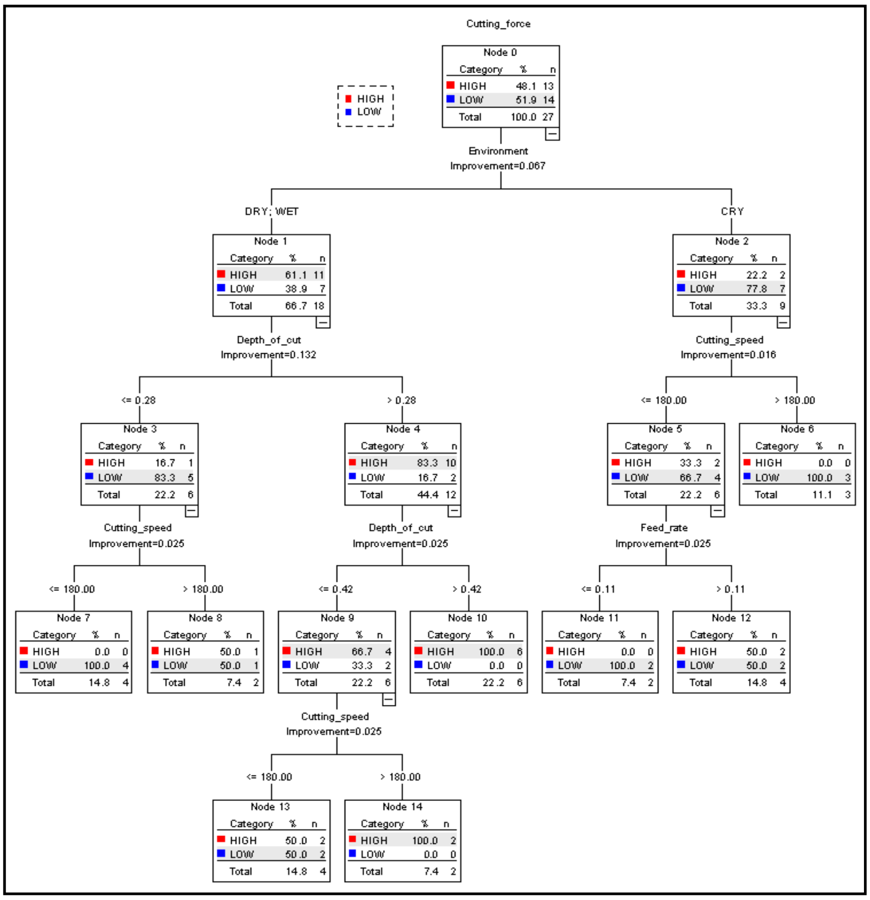


Figure 3.8 Classification tree for CF using CART algorithm

Rule 4: If Environment = Dry or wet, DOC ≤ 0.28 mm and CS ≤ 180 m/min Then CF is [106.23-171.30]

[P=100%, Q=28.57%, C=14.81%, QTY=4] [T=143.38%]

Rule 5: If Environment = Dry or wet, DOC ≤ 0.28 mm and CS > 180 m/min then CF is [106.23-171.30]

[P=100%, Q=7.69%, C=3.70%, QTY=1] [T=111.39%]

Rule 6: If Environment = Dry or wet, DOC > 0.28 mm Then CF is (171.30-286.90]

[P=100%, Q=46.15%, C=22.22%, QTY=6] [T=168.37%]

Rule 7: If Environment = Dry or wet, DOC > 0.28 mm and ≤ 0.42 mm, and CS ≤ 180 m/min Then CF is [106.23-171.30]

[P=50%, Q=14.28%, C=7.41%, QTY=2] [T=71.69%]

Rule 8: If Environment = Dry or wet, DOC > 0.28 mm and ≤ 0.42 mm, and CS >180 m/min Then CF is (171.30-286.90]

[P=100%, Q=15.38%, C=7.41%, QTY=2] [T=122.79%]

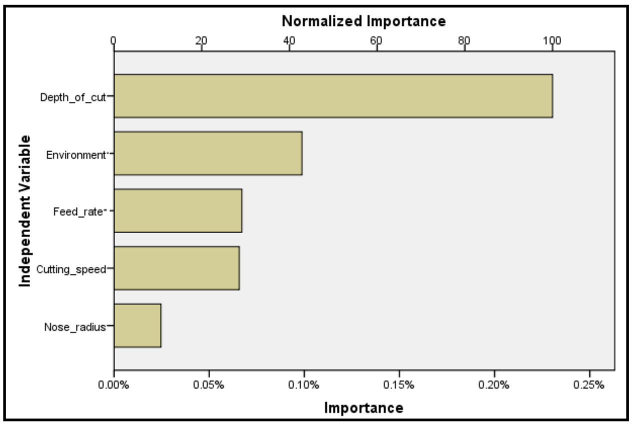


Figure 3.9 Importance of CNC turning parameters influencing CF

CHAID-based rules for cutting force:

Rule 1: If Environment = Cryogenic Then CF is [106.23-171.30]

[P=77.80%, Q=50%, C=25.92%, QTY=7] [T=153.72%]

Rule 2: If Environment = Dry or wet and DOC = Low Then CF is [106.23-171.30]

[P=83.30%, Q=35.71%, C=18.52%, QTY=5] [T=136.53%]

Rule 3: If Environment = Dry or wet, DOC = Medium or high and NR = Low or high Then CF is (171.30-286.90]

[P=100%, Q=61.54%, C=29.63%, QTY=8] [T=191.17%]

Rule 4: If Environment = Dry or wet, DOC = Medium or high and NR = Medium Then CF is [106.23-171.30]

[P=50%, Q=14.29%, C=7.41%, QTY=2] [T=71.70%]

In Table 3.7, a comparison of the classification accuracies for CART and CHAID algorithms for all the four responses is provided. From this table, it can be noted that for TL

Table 3.7 Classification accuracies for TL, PC, SR, and CF using CART and CHAID algorithms

Response	CART				CHAID			
TL		Predicted				Predicted		
	Observed	High (>27.66 min)	Low (≤ 27.66 min)	Percent correct	Observed	High (> 0.58 µm)	Low (≤ 0.58 µm)	Percent correct
	High (>27.66 min)	13	0	100%	High (>27.66 min)	12	1	92.3%
	Low (≤27.66 min)	0	14	100%	Low (≤27.66 min)	0	14	100%
	Overall percentage	48.1%	51.9 %	100%	Overall percentage	44.4%	55.6%	96.2%
PC		Predicted				Predicted		
	Observed	High (> 1373 W)	Low (≤ 1373 W)	Percent correct	Observed	High (> 1373 W)	Low(≤ 1373 W)	Percent correct
	High (> 1373 W)	12	0	100%	High (> 1373 W)	11	1	91.7%
	Low (≤ 1373W)	0	15	100%	Low (≤ 1373 W)	0	15	100%
	Overall percentage	44.4%	55.6 %	100%	Overall percentage	40.7%	59.3%	95.8%
SR		Predicted				Predicted		
	Observed	High (>0.58 µm)	Low (≤0.58 µm)	Percent correct	Observed	High (> 0.58 µm)	Low (≤ 0.58 µm)	Percent correct
	High (> 0.58 µm)	11	2	84.6 %	High (> 0.58 µm)	11	2	84.6%
	Low (≤ 0.58 µm)	3	11	78.6 %	Low (≤ 0.58 µm)	4	10	71.4%
	Overall percentage	51.8 %	48.1 %	81.6 %	Overall percentage	55.5%	44.4%	78.0%
CF		Predicted				Predicted		
	Observed	High (> 171.30 N)	Low (≤ 171.30 N)	Percent correct	Observed	High (> 0.58 µm)	Low (≤ 0.58µm)	Percent correct
	High (> 171.30 N)	12	1	85.7 %	High (>171.30 N)	8	5	61.5%
	Low (≤ 171.30 N)	4	10	76.9 %	Low (≤171.30 N)	0	14	100%
	Overall percentage	59.2 %	40.8 %	81.3 %	Overall percentage	29.6%	70.4%	80.7%

Table 3.8 Risk of classifying SR, TL, CF and PC

Response	Method	Accuracy	Standard error
TL	CART	1.000	0.001
	CHAID	0.963	0.036
PC	CART	1.000	0.001
	CHAID	0.958	0.036
SR	CART	0.815	0.075
	CHAID	0.778	0.080
CF	CART	0.818	0.075
	CHAID	0.807	0.036

and PC responses, CART algorithm can perfectly predict low and high TL, and low and high PC values. The classification accuracies for high and low SR are 84.6% and 78.6%, respectively. Similarly, using CART algorithm, high and low CFs can be predicted with accuracies of 85.7% and 76.9%, respectively. Thus, CART algorithm can almost perfectly predict both the high and low values of all the considered responses, although it has a slightly greater tendency to accurately estimate high values of the responses. In case of CHAID algorithm, it can perfectly predict lower values of TL, PC and CF. High values of TL and PC are predicted with 92.3% and 91.7% accuracies, respectively. It has prediction accuracies of 84.6% and 71.4% for high and low SR values, respectively. The classification accuracy for high CF is only 61.5%. Thus, it can be concluded that CHAID algorithm performs better in predicting lower values of the considered responses. The overall classification accuracies of both these algorithms for the four responses are provided in Table 3.8. With respect to the overall classification accuracy, CART algorithm outperforms CHAID in almost exactly predicting the responses for the considered CNC turning process. The corresponding values of standard error for CART are also comparatively low as compared to CHAID algorithm.

In order to visualize the effects of changing values of the responses of the considered CNC turning process on the prediction performance of both the CART and CHAID algorithms, a sensitivity analysis study is performed here. In this approach, incremental changes are made in the response values of the experimental dataset based on the following equation:

$$R_N = R_O + \left(2 \times RAND(\ ) - 1\right) \times e \times R_O \tag{3.1}$$

where R_O is the original response value, *RAND*() is a uniform random number generator between 0 and 1, e is the relative error level and R_N is the perturbed response value. The relative error levels are set here as 5, 10, 15, 20 and 25%. The prediction accuracies of both the classification algorithms at varying error levels are provided in Table 3.9. It can be clearly propounded that the prediction performance of CART algorithm is least influenced by the changing response values in the experimental data and it is a more robust technique as compared to CHAID algorithm.

Table 3.9 Classification accuracies of CART and CHAID algorithms at various error levels

Response	CART					CHAID				
	Error level									
	5%	10%	15%	20%	25%	5%	10%	15%	20%	25%
TL	100	100	100	92.6	92.6	92.6	92.6	92.6	82.6	85.2
PC	100	100	96.3	96.3	92.6	100	100	92.6	96.3	96.3
SR	88.9	81.5	85.2	85.2	81.5	81.5	80.9	80.45	81.5	79.0
CF	92.6	82.6	92.6	92.6	88.9	88.9	77.8	88.9	88.9	85.2

Since the past few decades, DT algorithms, like CART and CHAID, have been in extensive use for solving predictive analytics problems. As they are generic models based on an effective calculation procedure, they can easily arrive at the optimal solution for a given classification/prediction problem. DTs generated by those algorithms are efficient managerial tools that present all the decisions/outcomes in the form of a flowchart with branches and leaves. DTs solve problems of ML by transforming data into tree representation. Each branch of the tree symbolizes a decision option. The leaves at the end of the branches show the possible outcomes. DTs can deal with quantitative, qualitative or categorical attributes by assigning objects to a specific class in a classification problem. DT is one of the simplest and most popular classification algorithms to learn, understand and interpret. The DTs have several advantages, like requirement of less computational effort for data preparation during pre-processing, no need of normalization and scaling of data, least affected by the missing observation in the dataset, provision of explanation about a particular decision etc. Similarly, they also suffer from other disadvantages, like requirement of higher time for training, a small change in data may cause a large change in the tree structure causing instability, incapability to be applied for regression and prediction of continuous variables etc.

3.2 WEDM process

WEDM is a non-conventional material removal process where a continuously travelling electrically conductive wire is used as an electrode to erode material from a given workpiece. To explore its fullest machining potential, there is always a requirement to examine the effects of its varied input parameters on the responses and resolve the best parametric setting. In this example, parametric analysis of a WEDM process is carried out by applying non-parametric DT algorithms, based on a past experimental dataset. Two DT-based classification methods, i.e. CART and CHAID are considered here as the DM tools to examine the influences of six WEDM process parameters on the four responses, and identify the most preferred parametric mix to help in achieving the desired response values.

In a four-axis CNC WEDM set-up, Kumar et al. [124] performed 54 experiments to investigate the effects of six process parameters, e.g. pulse-on time (T_{on}), pulse-off time (T_{off}), peak current (I_P), spark gap voltage (SV), wire feed (WF) and wire tension (WT) on four responses/outputs, i.e. machining rate (MR) (in mm/min), SR (in μm), dimensional deviation

(DD) (in μm) and wire wear ratio (WWR). Other factors, like type of the workpiece material (pure titanium-grade 2), wire electrode (brass wire having 0.25 mm diameter), workpiece thickness and pressure of the dielectric were kept constant during the experimental runs. Each of the considered WEDM parameters had three distinct operating levels, as shown in Table 3.10. Table 3.11 exhibits the experimental plan and measured values of the responses. MR refers to the velocity at which the workpiece moves relative to the tool (or wire, in case of WEDM process). SR measures the quality of the machined surface, indicating how much the actual surface deviates from its ideal form in the normal direction. Large deviations result in a rough surface, while minimal deviations indicate a smooth surface. In WEDM process, DD is a measure of quality of the machined component because all of its dimensions must lie within the specified tolerances. It is quantified as the variation of actual dimension after machining from the target dimension. In this process, the profile traversed by the wire and the job profile are usually not similar. The distance between the actual profile and profile outlined by the wire is known as DD. The WWR is designated as the volume of material lost from the tool (wire) to the volume of material removed from the workpiece. Among these responses, MR is the sole beneficial quality characteristic desired with its higher value. On the contrary, minimum values are required for SR, DD and WWR (non-beneficial attributes) [125].

Table 3.10 WEDM process parameters with their levels [124]

Process parameter	Symbol	Level			Unit
		1	2	3	
Pulse-on time	T_{on}	112	116	120	μs
Peak current	I_P	120	160	200	A
Pulse-off time	T_{off}	44	50	56	μs
Spark gap voltage	SV	40	50	60	V
Wire tension	WT	500	950	1400	g
Wire feed	WF	4	7	10	m/min

Table 3.11 Experimental observations for the WEDM process [124]

Run No.	T_{on}	I_P	T_{off}	WF	SV	WT	MR	DD	SR	WWR
1	120	200	50	7	50	500	1.14	160	3.22	0.095
2	116	160	56	4	50	500	0.576	150	2.48	0.063
3	112	160	50	4	60	950	0.42	145	2.23	0.079
4	116	120	44	10	50	950	0.954	159	2.75	0.086
5	116	120	50	7	60	500	0.544	152	2.47	0.061
6	120	160	50	4	40	950	1.075	162	2.93	0.088
7	116	160	56	10	50	1400	0.586	150	2.48	0.063
8	116	160	50	7	50	950	0.695	152	2.65	0.080
9	116	160	44	4	50	500	1.014	160	2.81	0.089
10	120	160	50	10	40	950	1.075	160	2.94	0.088
11	120	160	56	7	40	950	0.995	160	2.91	0.087
12	120	160	50	4	60	950	0.809	159	2.83	0.079
13	116	160	44	10	50	500	1.012	160	2.79	0.076
14	116	160	50	7	50	950	0.573	150	2.61	0.064

Table 3.11 Contd.

15	112	120	50	7	50	500	0.406	145	2.49	0.048
16	116	160	50	7	50	950	0.697	152	2.68	0.082
17	116	120	50	7	60	1400	0.538	150	2.49	0.059
18	112	160	56	7	40	950	0.48	145	2.32	0.060
19	116	120	56	10	50	950	0.535	151	2.31	0.056
20	116	200	50	7	40	1400	0.825	152	2.89	0.079
21	116	200	50	7	60	500	0.773	152	2.69	0.072
22	116	200	56	10	50	950	0.792	153	2.57	0.074
23	116	120	50	7	40	1400	0.625	152	2.71	0.068
24	112	120	50	7	50	1400	0.425	145	2.51	0.054
25	116	200	56	4	50	950	0.799	155	2.56	0.078
26	120	160	50	10	60	950	0.81	153	2.82	0.081
27	120	120	50	7	50	500	0.83	158	2.77	0.074
28	112	160	50	10	40	950	0.521	150	2.35	0.085
29	112	200	50	7	50	500	0.535	150	2.48	0.083
30	112	160	44	7	40	950	0.858	153	2.70	0.089
31	112	200	50	7	50	1400	0.54	150	2.51	0.082
32	116	160	50	7	50	950	0.658	150	2.65	0.081
33	116	200	44	4	50	950	1.02	159	2.88	0.092
34	116	160	50	7	50	950	0.656	152	2.65	0.081
35	120	160	44	7	40	950	1.28	165	3.28	0.107
36	116	200	44	10	50	950	1.03	160	2.98	0.095
37	116	200	50	7	40	500	0.829	155	2.84	0.079
38	112	160	50	4	40	950	0.529	150	2.33	0.081
39	116	160	56	10	50	500	0.589	150	2.50	0.064
40	116	160	50	7	50	950	0.659	152	2.69	0.081
41	120	160	56	7	60	950	0.792	153	2.66	0.070
42	112	160	44	7	60	950	0.495	150	2.60	0.081
43	116	200	50	7	60	1400	0.778	155	2.68	0.072
44	116	120	44	4	50	950	0.959	155	2.75	0.086
45	112	160	50	10	60	950	0.429	145	2.28	0.079
46	120	120	50	7	50	1400	0.823	158	2.75	0.074
47	112	160	56	7	60	950	0.395	140	2.15	0.064
48	116	160	44	4	50	1400	0.981	159	2.85	0.088
49	116	120	50	7	40	500	0.635	158	2.78	0.068
50	120	160	44	7	60	950	1.00	159	3.00	0.085
51	116	120	56	4	50	950	0.541	150	2.29	0.060
52	120	200	50	7	50	1400	1.052	159	3.12	0.091
53	116	160	44	10	50	1400	0.962	155	2.82	0.088
54	116	160	56	4	50	1400	0.592	150	2.49	0.060
Minimum							0.40	140	2.15	0.05
Maximum							1.28	165	3.28	0.107
Median							0.74	152	2.68	0.079

In this work, CART and CHAID algorithms are applied to explore the experimental dataset of [124] while generating the corresponding DTs and induction rules to study the influences of the considered WEDM parameters on the four responses. An endeavour is also put forward to compare their relative classification performance with respect to prediction accuracy and prognosis risk. In a DT, an internal node denotes a test on an attribute, a branch characterizes the result of the test and a leaf node depicts a class label. A specific decision or induction rule can thus be obtained while following the path from the root to leaf node. For development of the related DTs using CART and CHAID algorithms available in SPSS 16.0, the following specifications are predefined.

For CART algorithm:

Growing method: CART, Categorical dependent variables: MR, SR, DD and WWR, Continuous independent variables: T_{on}, T_{off}, I_P, SV, WF and WT, Validation: Cross validation, Number of sample folds: 3, Growth limit: Maximum depth of tree = 5, Minimum number of cases: Parent node = 3, Child node = 2, Impurity measure: 'Gini', Minimum shift in improvement = 0.0001.

For CHAID algorithm:

Growing method: CHAID, Categorical dependent variables: MR, SR, DD and WWR, Categorical independent variables: T_{on}, T_{off}, I_P, SV, WF and WT, Validation: Cross validation, Number of sample folds: 3, Growth limit: Maximum depth of tree = 5, Minimum number of cases: Parent node = 3, Child node = 2, Significance level for: a) Splitting node = 0.03, b) Merging categories = 0.05 and c) Chi-square statistic = Pearson, Model estimation: a) Maximum number of iterations = 100, minimum variation in expected cell frequencies = 0.001 and c) Modify significance values using the Bonferroni method.

Figure 3.10 exhibits the CART algorithm-based DT, showing the impacts and contributions of the six WEDM parameters on MR. In this figure, the term 'Low' belongs to the observations with MR ≤ 0.74 mm/min and 'High' refers to those observations having MR > 0.74 mm/min (where 0.74 mm/min is the median value of MR). This classification tree contains 5 splits and 6 terminal nodes. The classification process begins with the top or root node with level 0. All the 54 observations are first assigned to this node where 27 observations are equally classified as 'Low' and 'High', as represented in this node. The root node is again divided into two new nodes and the related split condition is shown below the root node. It can be noticed from this DT that during the first classification, 42 experimental observations with T_{on} less than or equal to 118 µs are routed to node number 1, categorized as 'Low' items. The remaining 12 observations with T_{on} greater than 118 µs are directed to node number 2 with 'High' classification. All the 12 observations belonging to only 'High' value identify node 2 as a pure node with no misclassification error. In node 1, unequal number of observations recognizes it as an impure node providing more child nodes. Thus, node 1 is now split giving rise to two impure nodes. In this split, 10 observations with T_{off} less than or equal to 47 µs are dispatched to node number 3, and the rest 32 observations with T_{off} greater than 47 µs are directed to node number 4.

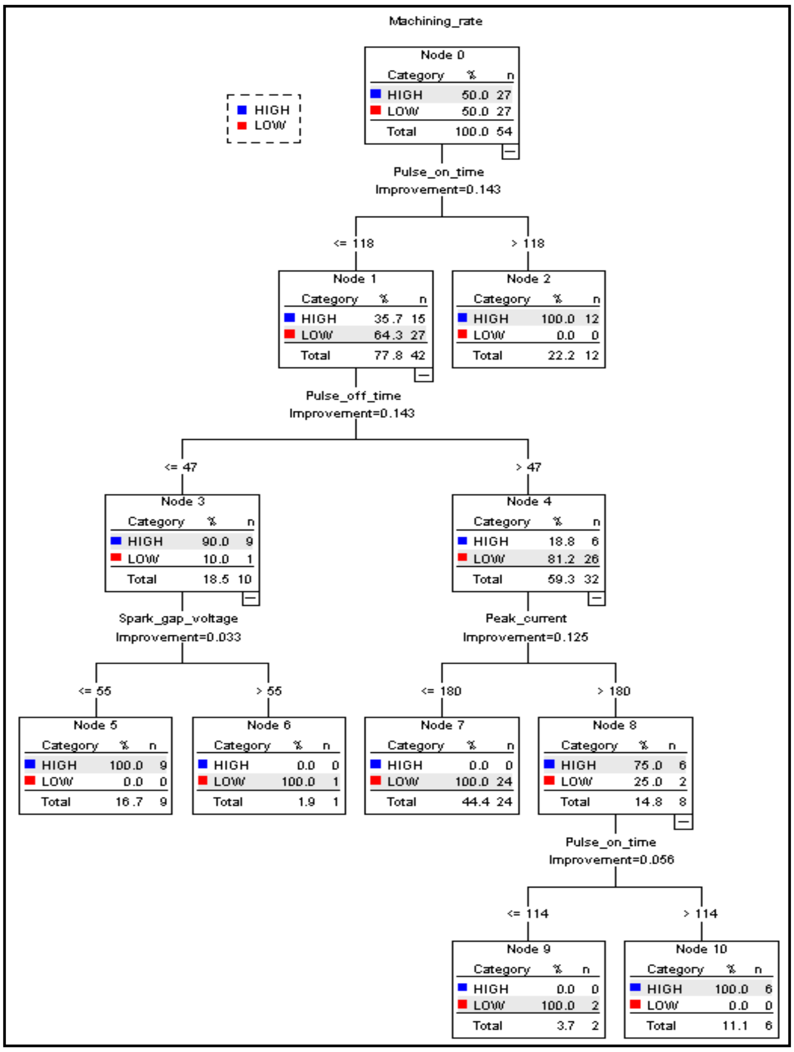


Figure 3.10 Decision tree showing the impacts of WEDM process parameters on MR

Now, based on the values of SV, node 3 is consequently divided into two pure nodes. In this split, nine observations with SV less than or equal to 55 V are attached to node number 5 with 'High' values of classification, and only one observation with SV greater than 55 V is routed to node number 6 with 'Low' classification value. Similarly, another splitting is carried out from node 4 based on I_P. In this split, 24 observations with I_P less than or equal to 180 A are dispatched to node number 7 with 'Low' categorization, and the remaining eight observations with I_P greater than 180 A are sent to node number 8. From node number 8, two pure nodes again emerge out based on the values of T_{on}. In this splitting operation, two observations having T_{on} less than or equal to 114 µs are routed to node number 9 and the rest six observations with T_{on} greater than 114 µs are sent to node number 10. In this DT, it is noticed that all the terminal nodes are pure representing no misclassification error. The percentage of correct classification at each of the terminal nodes is presented in Table 3.12. Now, from this tree, the following decision/induction rules can be structured to envisage the impacts of various WEDM process parameters on MR.

Table 3.12 Percentage of correct classification of MR based on CART algorithm

Classification	Terminal node	MR			
		Low (≤ 0.74 mm/min)		High (> 0.74 mm/min)	
		Number of observations	Percentage	Number of observations	Percentage
1	Node 2	0	0	12	100
2	Node 5	0	0	9	100
3	Node 6	1	100	0	0
4	Node 7	24	100	0	0
5	Node 9	2	100	0	0
6	Node 10	0	0	6	100

Decision rules for MR based on CART:

Rule 1: If T_{on} > 118 µs Then MR is (0.74-1.28]

[P=100%, Q=44.44%, C=22.22%, QTY=12, T=166.66%]

Rule 2: If T_{on} ≤ 118 µs and T_{off} ≤ 47 µs and SV≤ 55V Then MR is (0.74-1.28]

[P=100%, Q=33.33%, C=16.67%, QTY=9] [T=150%]

Rule 3: If T_{on} ≤ 118 µs and T_{off} ≤ 47 µs and SV> 55 V Then MR is [0.40-0.74]

[P=100%, Q=3.70%, C=1.85%, QTY=1] [T=105.55%]

Rule 4: If T_{on} ≤ 118 µs and T_{off} > 47 µs and I_P ≤ 180 A Then MR is [0.40-0.74]

[P=100%, Q=88.88%, C=44.44%, QTY=24] [T=233.32%]

Rule 5: If T_{on} ≤ 118 µs and T_{off} > 47 µs and I_P > 180 A and T_{on} ≤ 114 µs Then MR is [0.40-0.74]

[P=100%, Q=7.41%, C=3.70%, QTY=2] [T=111.11%]

Rule 6: If T_{on} ≤ 118 µs and T_{off} > 47 µs and I_P > 180 A and T_{on} > 114 µs Then MR is (0.74-1.28]

[P=100%, Q=22.22%, C=11.11%, QTY=6] [T=133.33%]

where P is the rule confidence, Q is the percentage of items in current equivalence class conforming to a rule, C is the rule support, QTY is the number of observations corresponding to a rule and T (total strength) = (P + Q + C) [126]. Among the developed decision rules, rule 4 with the maximum total strength of 233.32% states that in the considered WEDM process, when T_{on} is less than or equal to 118 µs, T_{off} is greater than 47 µs and I_P is less than or equal to 180 A, the corresponding MR would be low. On the other hand, rule 1 with a total strength of 166.67% reveals that high T_{on} leads to high MR. As MR is a beneficial response, it is thus advised to operate the WEDM process at high setting of T_{on} (above 118 µs). In the CART-based DT, where univariate splits are considered, the predictor variables are usually rated on a 0-100 scale depending on their preference in accounting for the responses on the dependent variable (MR). It can be noticed from Fig. 3.11 that T_{on} has the maximum influence on MR response, followed by T_{off}, I_P and SV. In this WEDM process, MR is observed to be totally unaffected by WF and WT.

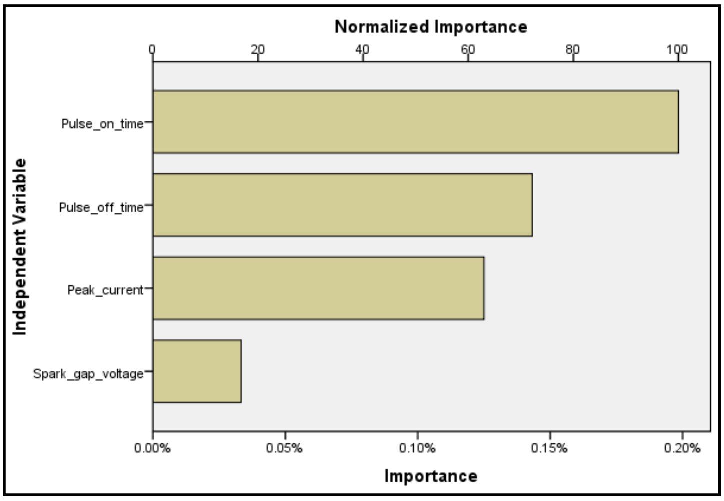


Figure 3.11 Importance of various WEDM parameters on MR

Similarly, the developed DT for MR using CHAID algorithm is exhibited in Figure 3.12. As compared to CART algorithm, same number of classifications is also obtained in CHAID algorithm. In this multi-split DT generated using CHAID, node 7 contains non-homogenous observations, identifying it as an impure node with 1.85% misclassification error. The corresponding decision rules identify T_{on}, T_{off} and I_P as the most indicative WEDM parameters influencing MR, whereas, SV, WF and WT appear to be insignificant parameters.

Decision rules for MR based on CHAID:

Rule 1: If T_{on} = High Then MR is (0.74-1.28]

[P=100%, Q=44.44%, C=22.22%, QTY=12] [T=166.66%]

Rule 2: If T_{on} = Medium and T_{off} = Low Then MR is (0.74-1.28]

[P=100%, Q=29.62%, C=14.81%, QTY=8] [T=144.43%]

Rule 3: If T_{on} = Low and T_{off} = Low Then MR is (0.74-1.28]

[P=100%, Q=37.04%, C=18.52%, QTY=10] [T=155.56%]

Rule 4: If T_{on} = Low and T_{off} = Medium or High Then MR is [0.40-0.74]

[P=50%, Q=3.70%, C=1.85%, QTY=1] [T=55.55%]

Rule 5: If T_{on} = Medium and T_{off} = Medium or High and I_P = High Then MR is (0.74-1.28]

[P=100%, Q=22.22%, C=11.11%, QTY=6] [T=133.33%]

Rule 6: If T_{on} = Medium and T_{off} = Medium or High and I_P = Medium or Low Then MR is [0.40-0.74]

[P=100%, Q=59.26%, C=29.63%, QTY=16] [T=188.89%]

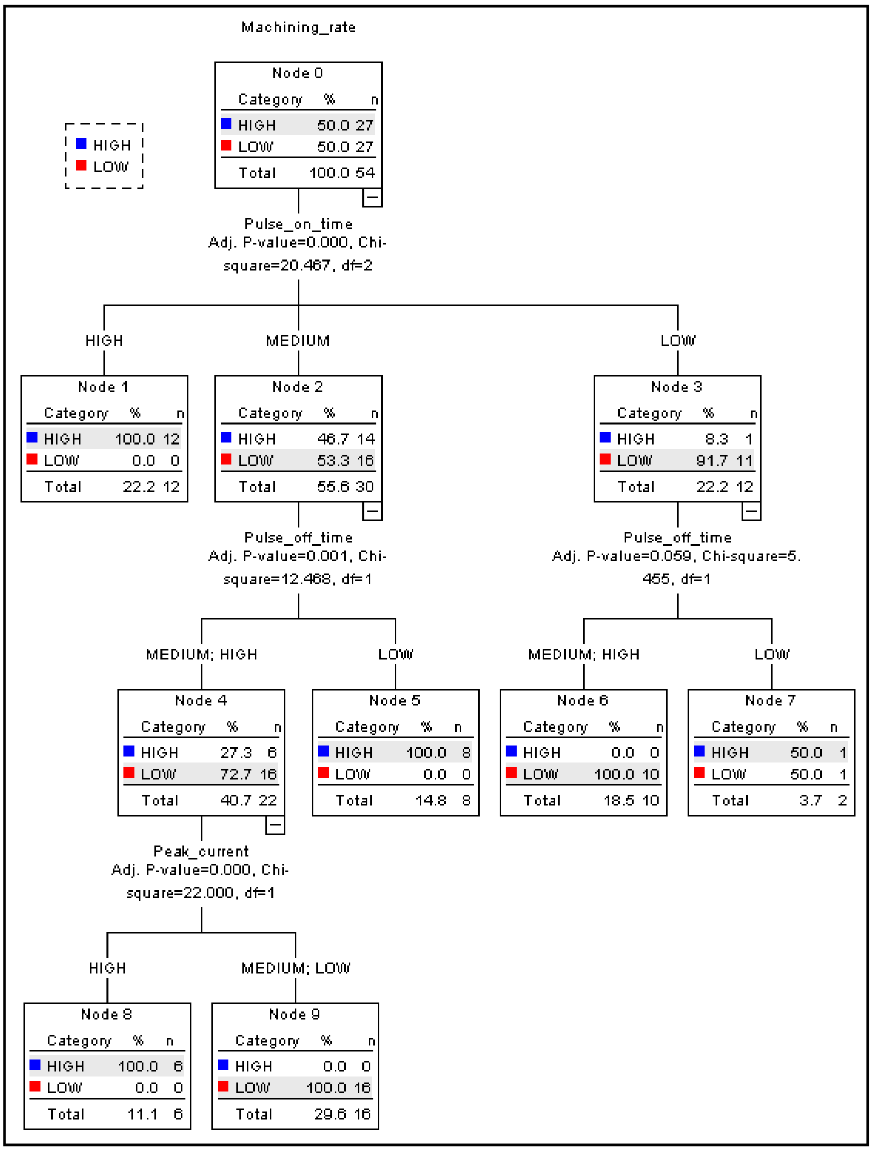


Figure 3.12 CHAID algorithm-based DT for MR

The CART algorithm-based DT representing the influences of varied WEDM parameters on SR is depicted in Figure 3.13. In this case, the measured SR values are also assorted into two classes, i.e. ‘Low’ containing SR values ≤ 2.68 μm and ‘High’ has SR values > 2.68 μm (where 2.68 μm is the median value obtained from 54 experimental runs). As it is a non-beneficial response, its low values are always preferred. The induction rules formulated from the developed DT state that when T_{on} is greater than 118 μs, the resulting SR would be high (rule 1 with total strength of 154.37%). Rule 4 with the maximum strength of 216.74% reveals that when T_{on} is less than or equal to 118 μs, T_{off} is greater than 47 μs and SV is greater than 45 V, SR would be low. The rules extracted from the DT developed using CHAID algorithm, as exhibited in Figure 3.14, also prove that high T_{on} results in high SR. Medium or low T_{off} is also responsible for increased SR. Similarly, low SV is responsible for high value of SR. In both the sets of rules developed using CART and CHAID algorithms, WF and WT appear to be irrelevant WEDM process parameters having no influences on SR.

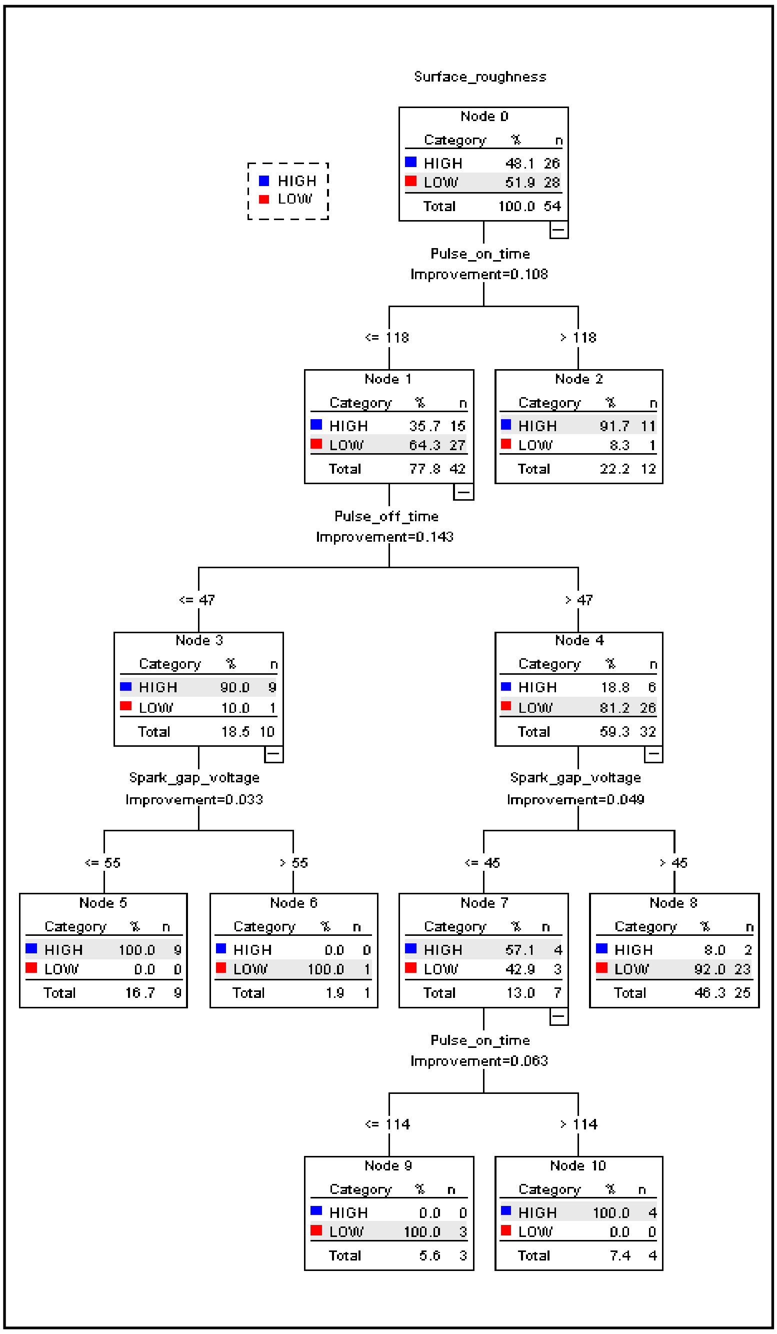


Figure 3.13 Decision tree for SR using CART algorithm

Figure 3.15 depicts the relative importance of various WEDM process parameters on SR. It can be revealed from this figure that T_{on} plays the dominant role in controlling SR of the machined components, followed by T_{off} and SV. The other three process parameters, i.e. I_P, WF and WT have less importance on SR.

Decision rules for SR using CART:

Rule 1: If $T_{on} > 118$ µs Then SR is (2.68-3.28)

[P=91.70%, Q=42.30%, C=20.37%, QTY=11] [T=154.37%]

Rule 2: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV≤ 55V Then SR is (2.68-3.28]

[P=100%, Q=34.61%, C=16.67%, QTY=9] [T=151.28%]

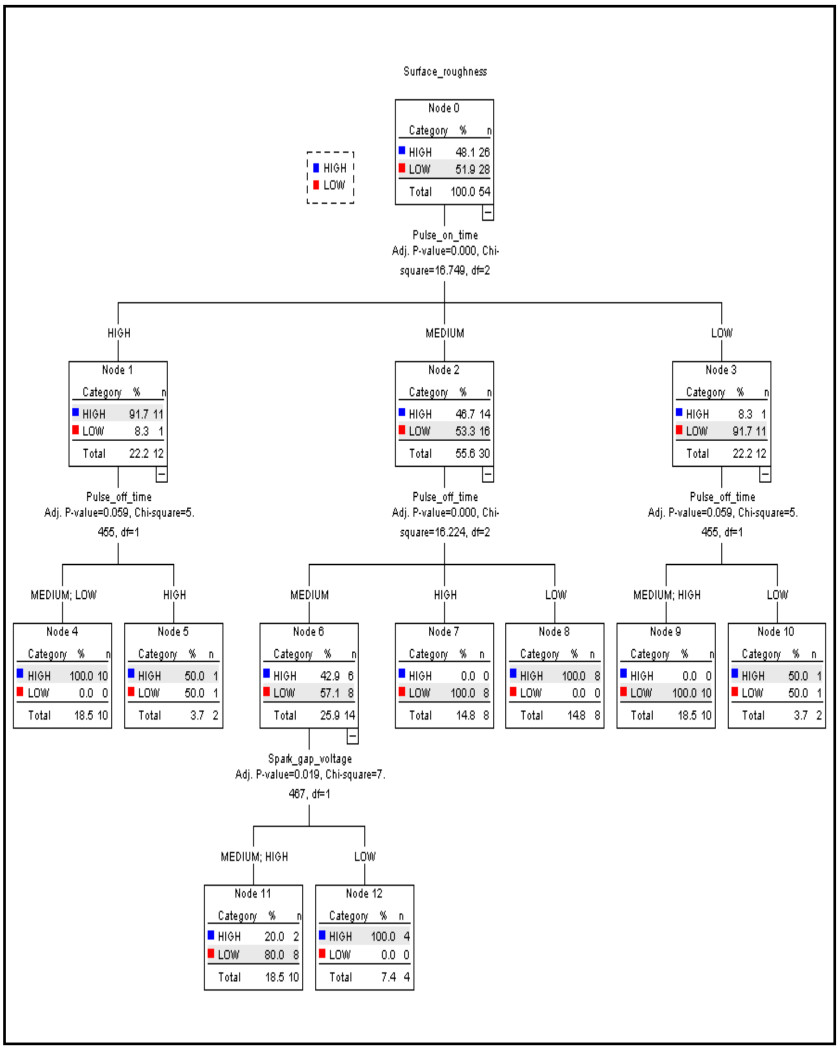


Figure 3.14 CHAID algorithm-based DT for SR

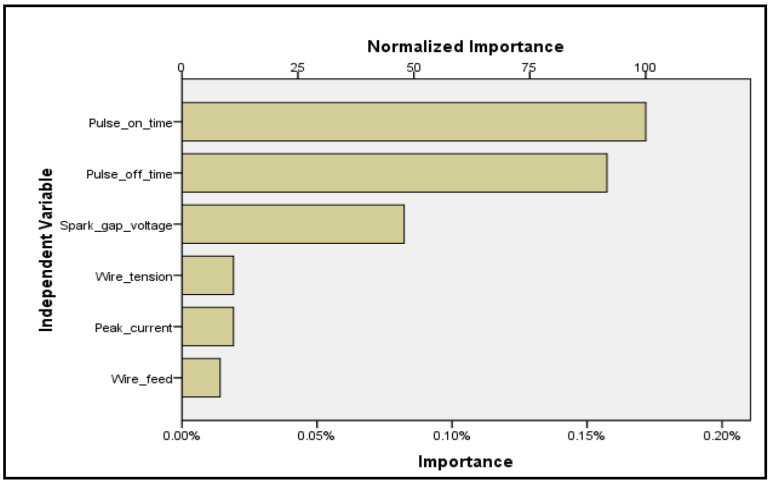


Figure 3.15 Importance of various WEDM parameters on SR

Rule 3: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV> 55V Then SR is [2.15-2.68] [P=100%, Q=3.57%, C=1.85%, QTY=1] [T=105.42%]

Rule 4: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and SV > 45V Then SR is [2.15-2.68] [P=92%, Q=82.14%, C=42.60%, QTY=23] [T=216.74%]

Rule 5: If $T_{on} \leq 118$ μs and $T_{off} > 47$ μs and SV ≤ 45V and $T_{on} \leq 114$ μs Then SR is [2.15-2.68]

[P=100%, Q=10.71%, C=5.55%, QTY=3] [T=116.26%]

Rule 6: If $T_{on} \leq 118$ μs and $T_{off} > 47$ μs and SV ≤ 45V and $T_{on} > 114$ μs Then SR is (2.68-3.28]

[P=100%, Q=15.38%, C=7.42%, QTY=4] [T=122.80%]

Decision rules for SR using CHAID:

Rule 1: If T_{on} = High and T_{off} = Medium or Low Then SR is (2.68-3.28]

[P=100%, Q=38.46%, C=18.51%, QTY=10] [T=156.97%]

Rule 2: If T_{on} = High and T_{off} = High Then SR is (2.68-3.28]

[P=500%, Q=3.85%, C=1.85%, QTY=1] [T=55.70%]

Rule 3: If T_{on} =Medium and T_{off} = High Then SR is [2.15-2.68]

[P=100%, Q=28.57%, C=14.81%, QTY=8] [T=143.38%]

Rule 4: If T_{on} = Medium and T_{off} = Low Then SR is (2.68-3.28]

[P=100.00%, Q=30.77%, C=14.81%, QTY=8] [T=145.58%]

Rule 5: If T_{on} = Low and T_{off} = Medium or High Then SR is [2.15-2.68]

[P=100%, Q=35.71%, C=18.51%, QTY=10] [T=155.22%]

Rule 6: If T_{on} = Low and T_{off} = Low Then SR is (2.68-3.28]

[P=50%, Q=3.84%, C=1.86%, QTY=1] [T=55.70%]

Rule 7: If T_{on} = Medium and T_{off} = Medium and SV = Medium or High Then SR is [2.15-2.68]

[P=80%, Q=28.57%, C=14.81%, QTY=8] [T=123.38%]

Rule 8: If T_{on} = Medium and T_{off} = Medium and SV= Low Then SR is (2.68-3.28]

[P=100%, Q=15.38%, C=7.41%, QTY=4] [T=122.79%]

To visualize the effects of various WEDM parameters on DD response, the corresponding DT of Figure 3.16 is now developed using CART algorithm. When the values of DD are less than or equal to 152 μm, they are denoted as 'Low' and when its values are greater than 152 μm, they are designated as 'High' (where 152 μm is the median of DD for the given set of observations). The 'If-Then' rules extracted from this DT highlight that T_{on} less than or equal to 118 μs, T_{off} greater than 47 μs and I_P less than or equal to 180 A always lead to lower DD (rule 4 with total strength of 230.54%). Higher T_{on} is liable for higher DD (rule 1 with total strength of 168.37%). Low T_{on}, high T_{off} and low SV would cause lower DD. The induction rules generated based on CHAID algorithm for DD are portrayed in Figure 3.17. Combining both the sets of decision rules from CART and CHAID algorithms, the optimal parametric mix of low T_{on}, medium or high T_{off}, low I_P and high SV would always lead to lower DD.

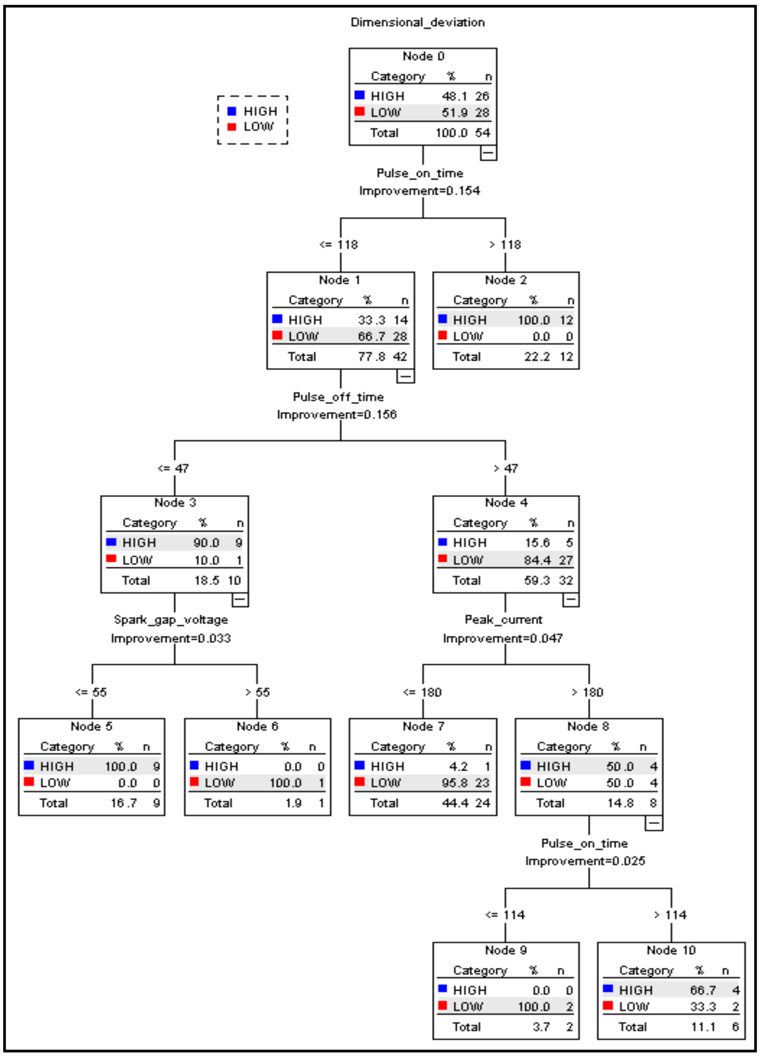


Figure 3.16 Decision tree for DD using CART algorithm

Kumar et al. [124] also observed that almost the same combination of the input parameters would be responsible for attaining lower DD in the said WEDM process. The importance plot of Figure 3.18 identifies T_{on} as the most critical WEDM parameter affecting DD, followed by T_{off}, I_P and SV. Interestingly, WF has no significant role in controlling the DD of the machined components.

Decision rules for DD using CART:

Rule 1: If $T_{on} > 118$ µs Then DD is (152-165]

[P=100%, Q=46.15%, C=22.22%, QTY=12] [T=168.37%]

Rule 2: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV≤ 55V Then DD is (152-165]

[P=100%, Q=34.61%, C=16.67%, QTY=9] [T=151.28%]

Rule 3: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV > 55V Then DD is [140-152]

[P=100%, Q=3.57%, C=1.85%, QTY=1] [T=105.42%]

Rule 4: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P \leq 180$A Then DD is [140-152]

[P=95.80%, Q=82.14%, C=52.60%, QTY=23] [T=230.54%]

Rule 5: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and I_P>180A and $T_{on} > 114$ µs Then DD is (152-165] [P=66.67%, Q=15.38%, C=7.41%, QTY=4] [T=89.46%]

Rule 6: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and I_P>180A and $T_{on} \leq 114$ µs Then DD is [140-152] [P=100%, Q=7.14%, C=3.70%, QTY=2] [T=110.84%]

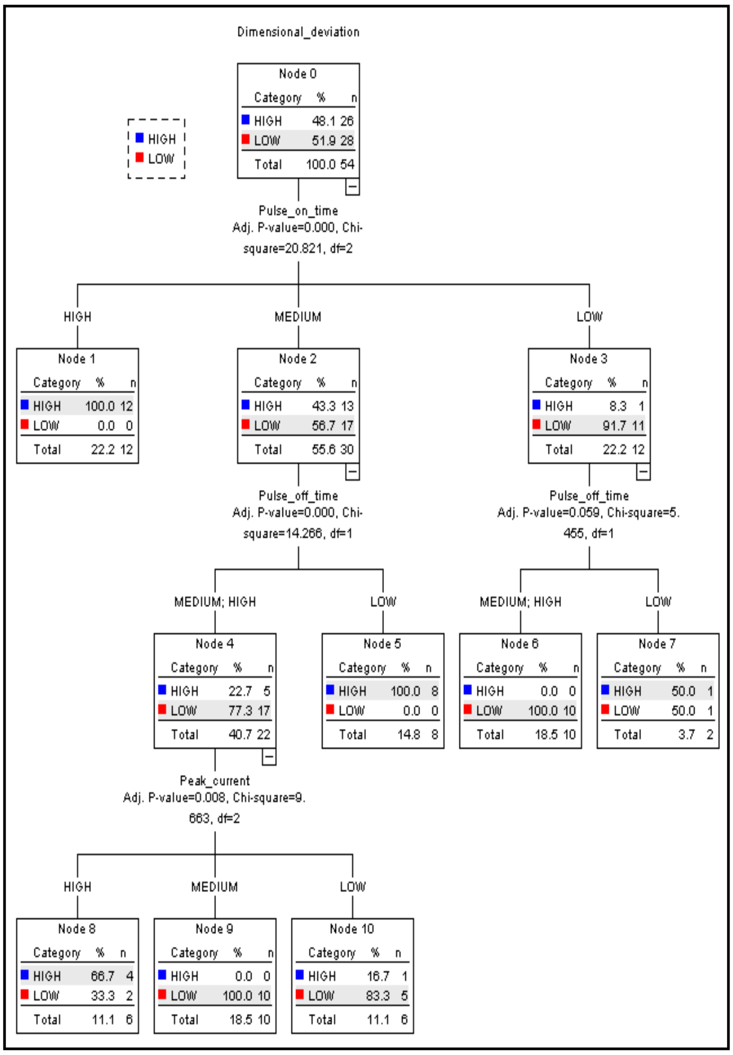


Figure 3.17 CHAID algorithm-based DT for DD

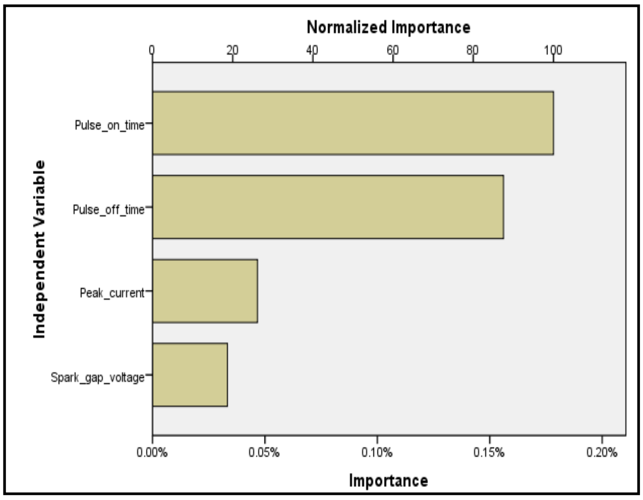


Figure 3.18 Importance of varied WEDM parameters on DD

The DT for WWR originated using CART algorithm is shown in Figure 3.19. The corresponding ‘If-Then’ rules are subsequently generated from it. When the values of WWR are less than or equal to 0.079, they are denoted as ‘Low’ and when its values are greater than 0.079, they are designated as ‘High’ (where 0.079 is the median of WWR estimated from the experimental dataset). An analysis of these rules reveals that when T_{off} is greater than 47 μs and I_P is less than or equal to 140 A, the achievable WWR is low (rule 2 with total strength of 154.23%). Similarly, lower T_{off} leads to higher WWR (rule 1 having total strength of 154.38%). The rules extracted from the DT based on CHAID algorithm, as shown in Figure 3.20, state that T_{off} would considerably affect WWR. On the contrary, WT is least responsible for attaining lower WWR.

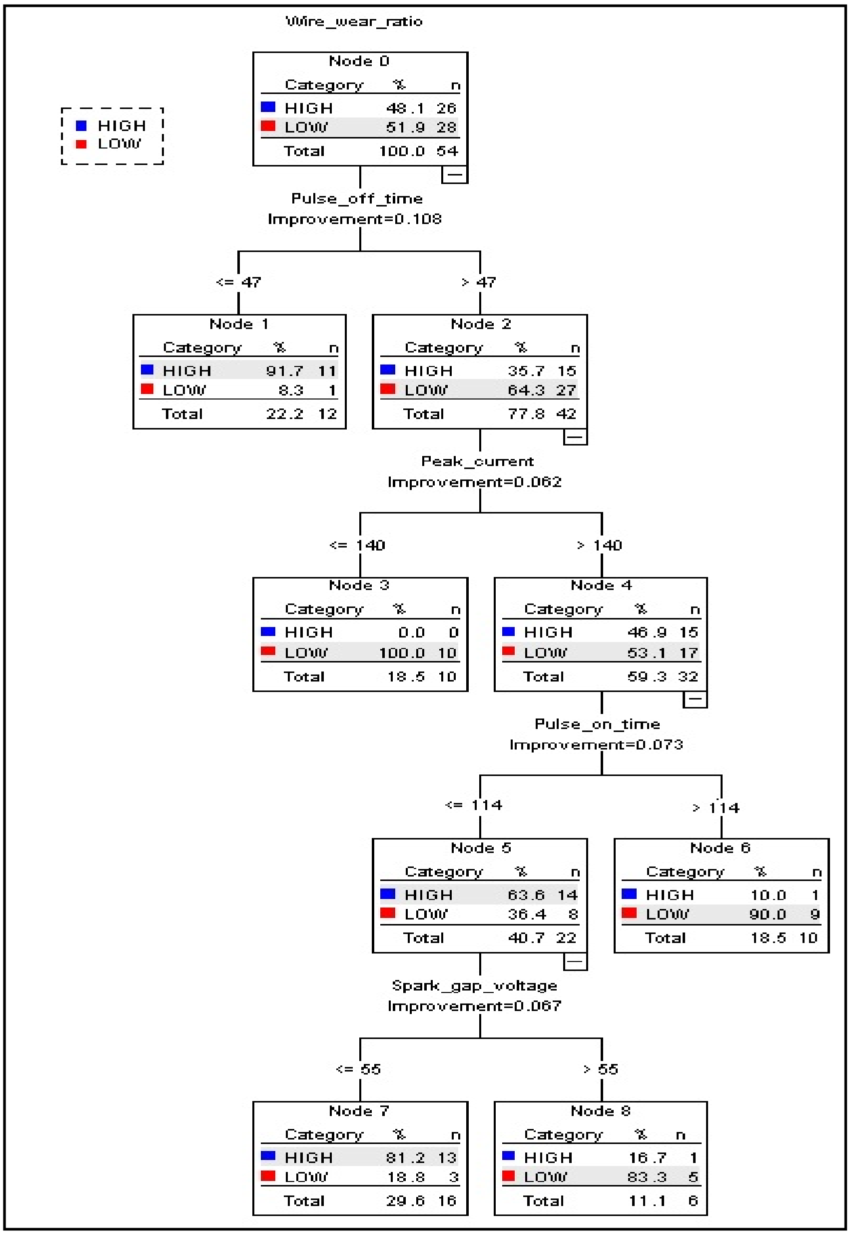


Figure 3.19 Decision tree for WWR using CART algorithm

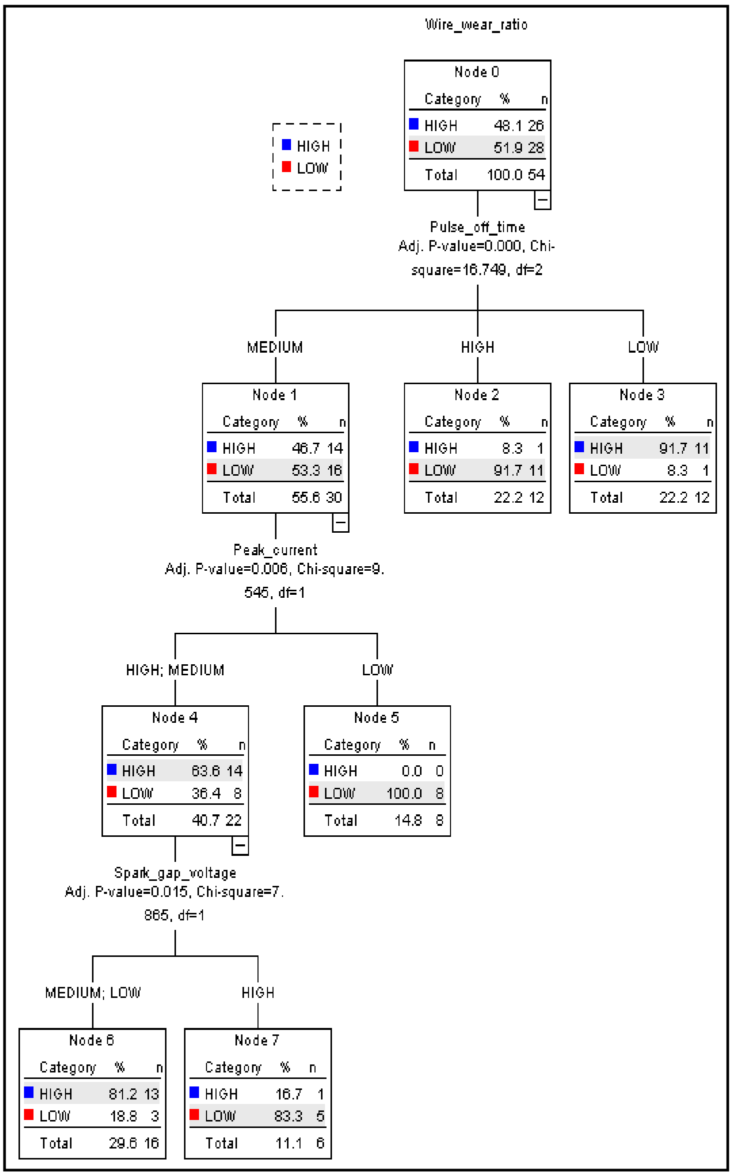


Figure 3.20 CHAID algorithm-based decision tree for WWR

From Figure 3.21, T_{off} is identified as the most indicative parameter influencing WWR, followed by SV, I_P and T_{on}. WF and WT appear to be insignificant WEDM parameters for WWR. An optimal parametric mix of low or moderate T_{off}, high I_P, low or moderate T_{on} and moderate SV had been identified by Kumar et al. [124] for attaining lower WWR, which almost corroborate with the DT-based observations.

Decision rules for WWR using CART:

Rule 1: If $T_{off} \leq 47$ µs Then WWR is (0.079-0.107]

[P=91.70%, Q=42.31%, C=20.37%, QTY=11] [T=154.38%]

Rule 2: If $T_{off} > 47$ µs and $I_P \leq 140$ A Then WWR is [0.05-0.079]

[P=100%, Q=35.71%, C=18.52%, QTY=10] [T=154.23%]

Rule 3: If $T_{off} > 47$ µs and $I_P > 140$ A and $T_{on} > 114$ µs Then WWR is [0.05-0.079]

[P=90%, Q=32.14%, C=16.67%, QTY=9] [T=138.81%]

Rule 4: If T_{off} > 47 µs and I_P>140 A and T_{on}≤114 µs and SV≤55V Then WWR is (0.079-0.107]

[P=81%, Q=50.00%, C=24.07%, QTY=13] [T=155.07%]

Rule 5: If T_{off} > 47 µs and I_P>140 A and T_{on}≤114 µs and SV>55V Then WWR is [0.05-0.079]

[P=83.30%, Q=17.86%, C=9.26%, QTY=5] [T=110.42%]

Decision rules for WWR based on CHAID:

Rule 1: If T_{off} = High Then WWR is [0.05-0.079]

[P=91%, Q=39.29%, C=20.37%, QTY=11] [T=150.66%]

Rule 2: If T_{off} = Low Then WWR is (0.079-0.107]

[P=91%, Q=42.31%, C=20.37%, QTY=11] [T=153.68%]

Rule 3: If T_{off} = Medium and I_P= Low Then WWR is [0.05-0.079]

[P=100%, Q=28.57%, C=14.81%, QTY=8] [T=143.38%]

Rule 4: If T_{off} = Medium and I_P= High or Medium and SV=Medium or Low Then WWR is (0.079-0.107]

[P=81.02%, Q=50%, C=24.07%, QTY=13] [T=155.09%]

Rule 5: If T_{off} = Medium and I_P= High or Medium and SV=High Then WWR is [0.05-0.079]

[P=83.30%, Q=17.86%, C=9.26%, QTY=5] [T=110.42%]

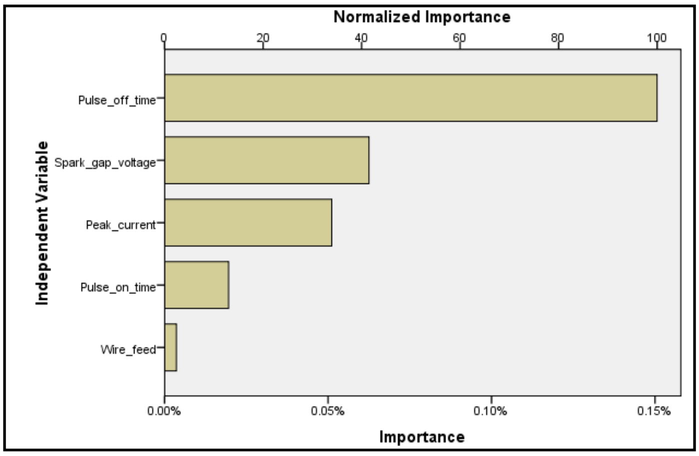


Figure 3.21 Importance of WEDM process parameters on WWR

In Table 3.13, a comparison of the classification accuracies between CART and CHAID algorithms for all the four WEDM responses is provided. From this table, it can be noted that for MR response, CART algorithm can perfectly predict both of its low and high values. For this algorithm, the classification accuracies for high and low SR values are 92.3% and 96.4%, respectively. Similarly, using CART algorithm, high and low DD values can be predicted with accuracies of 96.2% and 92.9%, respectively. For WWR response, CART algorithm can respectively predict the corresponding high and low values with 92.30% and 89.28% accuracies. Thus, CART algorithm can almost perfectly predict both the high and low values of all the WEDM responses, although it has a slightly greater likelihood to accurately

Table 3.13 Classification accuracies for MR, SR, DD and WWR using CART and CHAID algorithms

Response	CART				CHAID			
MR		Predicted				Predicted		
	Observed	High (> 0.74 mm/min)	Low (≤ 0.74 mm/min)	Percent correct	Observed	High (> 0.74 mm/min)	Low (≤ 0.74 mm/min)	Percent correct
	High (> 0.74 mm/min	27	0	100	High (> 0.74 mm/min)	27	0	100
	Low (≤ 0.74 mm/min)	0	27	100	Low (≤ 0.74 mm/min)	1	26	96.3
	Overall percentage	50	50	100	Overall percentage	51.9	48.1	98.1
SR		Predicted				Predicted		
	Observed	High (> 2.68 μm)	Low (≤ 2.68 μm)	Percent correct	Observed	High (> 2.68 μm)	Low (≤ 2.68 μm)	Percent correct
	High (> 2.68 μm)	24	2	92.3	High (> 2.68 μm)	24	2	92.3
	Low (≤ 2.68 μm)	1	27	96.4	Low (≤ 2.68 μm)	2	26	92.9
	Overall percentage	46.3	53.7	94.4	Overall percentage	48.1	51.9	92.6
DD		Predicted				Predicted		
	Observed	High(> 152 μm)	Low(≤ 152 μm)	Percent correct	Observed	High (> 152 μm)	Low (≤ 152 μm)	Percent correct
	High (> 152 μm)	25	1	96.2	High (> 152 μm)	25	1	96.2
	Low (≤ 152 μm)	2	26	92.9	Low (≤ 152 μm)	3	25	89.3
	Overall percentage	50	50	94.5	Overall percentage	51.9	48.1	92.6
WWR		Predicted				Predicted		
	Observed	High (> 0.079)	Low (≤ 0.079)	Percent correct	Observed	High (> 0.079)	Low (≤ 0.079)	Percent correct
	High (> 0.079)	24	2	92.30	High (> 0.079)	24	2	92.3
	Low (≤ 0.079)	3	25	89.28	Low (≤ 0.079)	4	24	85.7
	Overall percentage	50	50	90.80	Overall percentage	51.9	48.1	88.9

Table 3.14 Risk of classifying MR, SR, DD and WWR

Response	Method	Accuracy	Standard error
MR	CART	1.000	0.001
	CHAID	0.981	0.018
SR	CART	0.944	0.032
	CHAID	0.926	0.036
DD	CART	0.945	0.031
	CHAID	0.926	0.036
WWR	CART	0.908	0.043
	CHAID	0.889	0.046

envisage higher values of the responses. In case of CHAID algorithm, high values of MR and SR are respectively predicted with 100% and 92.3% accuracies. It has prediction accuracies of 96.2% and 89.3%, respectively for high and low DD values. For this algorithm, the classification accuracy for low WWR is the least (85.7%). The overall classification accuracies of both these algorithms for the four responses are also provided in Table 3.13. With respect to this measure, CART algorithm outperforms CHAID algorithm for all the responses. The corresponding standard errors of the prediction accuracies for CART algorithm are also comparatively low against CHAID algorithm, as provided in Table 3.14. It confirms superiority of CART algorithm in almost exactly anticipating all the response values for the considered WEDM process.

3.3 USM process

Kataria et al. [127] conducted 36 experiments to study the influences of various experimental parameters, such as cobalt content (CC), thickness of the workpiece (TW), abrasive grit size (GS), tool profile (TP), tool material and power rating (PR) on the responses, like MRR and TWR during ultrasonic drilling of WC-Co composite material. Each of those process parameters was set at different levels, i.e. CC (6%, 24%), TW (3 mm, 5 mm), TP (solid, hollow), tool material (stainless steel, silver steel, nimonic-80A), GS (200, 320, 500) and PR (40%, 60%, 80%). Table 3.15 represents the detailed experimental design plan along with the measured responses, which is later subjected to CART analysis to identify the most influencing parameters by developing the corresponding DTs with the following specifications:

a) Split selection method - CART style exhaustive search for univariate splits,

b) Measure of goodness of fit - Gini index, Prior probabilities - estimated, misclassification cost - equal,

c) Stopping rule - Prune on misclassification error,

d) Stopping parameters - minimum n = 5,

e) Standard error rule - 1.0,

f) Sampling - seed for random number generator = 12,

g) V-fold cross-validation = 3.

Table 3.15 Experimental design plan and measured responses for the USM process [127]

Sl. No.	USM parameter						Response	
	CC (%)	TW (mm)	TP	Tool material	GS (mesh no.)	PR (%)	MRR (g/min)	TWR (g/min)
1	6	3	Solid	Stainless Steel	200	40	0.0112	0.0041
2	6	3	Solid	Stainless Steel	320	60	0.0126	0.0030
3	6	3	Solid	Stainless Steel	500	80	0.0124	0.0029
4	6	5	Hollow	Stainless Steel	200	40	0.0058	0.0029
5	6	5	Hollow	Stainless Steel	320	60	0.0163	0.0040
6	6	5	Hollow	Stainless Steel	500	80	0.0138	0.0038
7	24	3	Hollow	Stainless Steel	200	40	0.0071	0.0023
8	24	3	Hollow	Stainless Steel	320	60	0.0170	0.0046
9	24	3	Hollow	Stainless Steel	500	80	0.0140	0.0033
10	24	5	Solid	Stainless Steel	200	40	0.0055	0.0022
11	24	5	Solid	Stainless Steel	320	60	0.0071	0.0019
12	24	5	Solid	Stainless Steel	500	80	0.0071	0.0016
13	6	3	Solid	Silver Steel	200	60	0.0171	0.0037
14	6	3	Solid	Silver Steel	320	80	0.0248	0.0050
15	6	3	Solid	Silver Steel	500	40	0.0032	0.0018
16	6	5	Hollow	Silver Steel	200	60	0.0210	0.0041
17	6	5	Hollow	Silver Steel	320	80	0.0404	0.0062
18	6	5	Hollow	Silver Steel	500	40	0.0046	0.0019
19	24	3	Hollow	Silver Steel	200	60	0.0199	0.0045
20	24	3	Hollow	Silver Steel	320	80	0.0295	0.0055
21	24	3	Hollow	Silver Steel	500	40	0.0031	0.0014
22	24	5	Solid	Silver Steel	200	60	0.0093	0.0028
23	24	5	Solid	Silver Steel	320	80	0.0132	0.0027
24	24	5	Solid	Silver Steel	500	40	0.0027	0.0013
25	6	3	Solid	Nimonic-80A	200	80	0.0423	0.0107
26	6	3	Solid	Nimonic-80A	320	40	0.0111	0.0026
27	6	3	Solid	Nimonic-80A	500	60	0.0056	0.0021
28	6	5	Hollow	Nimonic-80A	200	80	0.0396	0.0073
29	6	5	Hollow	Nimonic-80A	320	40	0.0070	0.0019
30	6	5	Hollow	Nimonic-80A	500	60	0.0173	0.0046
31	24	3	Hollow	Nimonic-80A	200	80	0.0436	0.0111
32	24	3	Hollow	Nimonic-80A	320	40	0.0113	0.0024
33	24	3	Hollow	Nimonic-80A	500	60	0.0070	0.0022
34	24	5	Solid	Nimonic-80A	200	80	0.0275	0.0083
35	24	5	Solid	Nimonic-80A	320	40	0.0067	0.0025
36	24	5	Solid	Nimonic-80A	500	60	0.0040	0.0015
					Average		0.0150	0.0037

The DM tool in the form of CART algorithm is applied to analyze the influences of various USM parameters on the responses, i.e. MRR and TWR, in the form of DTs and induction of decision rules. The DT for MRR, as depicted in Figure 3.22, is developed to analyze the effect of different machining parameters on MRR values. In the developed tree, 'Class 1' label consists of MRR values ≤ 0.0150 g/min and 'Class 2' label denotes MRR values > 0.0150 g/min, where 0.0150 is the average value of MRR as obtained from the experimental data. The developed tree has 6 splits and 7 terminal nodes. Terminal nodes (leaves), shown in red outlines, are those points on the tree where no further classification is possible, i.e. all the elements in the node belong to the same class. While rest of the decision nodes are outlined in blue lines. The DT begins with a single node, i.e. the top root node, which is represented as node 1. At first, all the 36 datasets are allotted to the root node and grouped as Class 1, as indicated by Class 1 label at the top extreme right of this node. In the initial stage of classification, 'Class 1' is selected because there are more Class 1 datasets as represented by the corresponding histogram shown inside the node. The root node then generates two new child nodes with PR as the best split variable. The data sets with PR values lower than or equal to 50% are routed to the second node and assigned as Class 1 label, and those with values greater than 50% are assigned to node number 3 and marked with Class 2 label. The attribute values 12 and 24 described above the nodes 2 and 3, respectively, signify the number of cases sent to each of these child nodes from their internal decision node. At node 2, histograms with same length above and below the reference line recognize it as a terminal node where no further partitioning or classification can be possible. On the other hand, node 3 acts as an impure node which can further generate new child nodes due to unequal length of the histograms displayed in the same. Thus, from node 3, 16 datasets with GS values less than or equal to 410 (mesh) are classified as Class 2 at node number 4, whereas, node number 5 is grouped as Class 1 classification with remaining 8 datasets having GS greater than 410 (mesh).

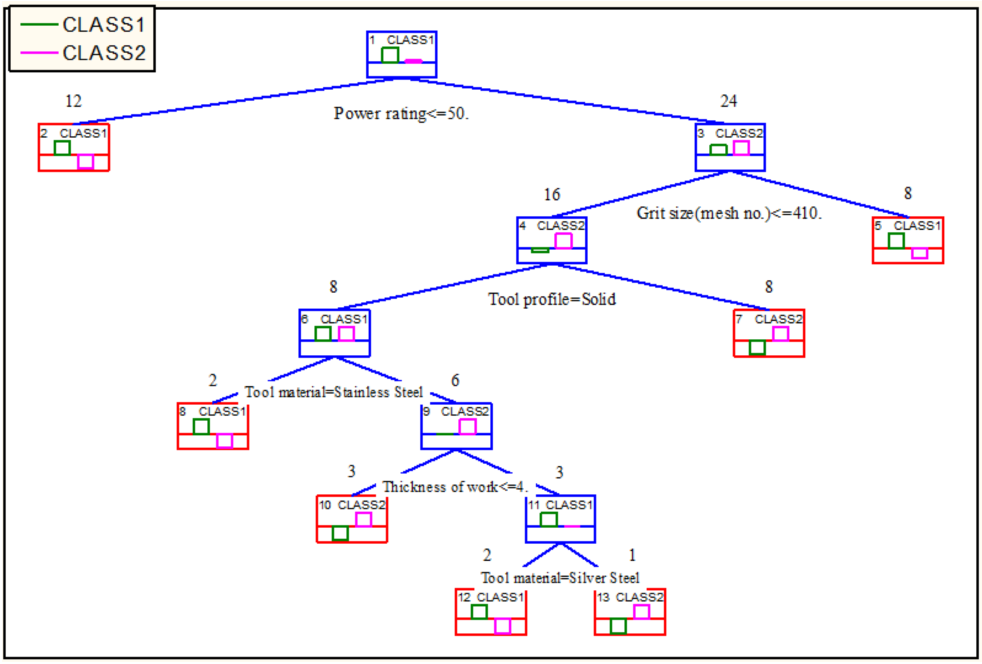


Figure 3.22 CART-based DT for MRR

Next, based on the TP attribute = solid, node 4 is further split into two nodes with 8 datasets labelled as Class 2 in node number 7 which can be identified as a pure node, while 8 datasets are assigned to node 6 and categorized as Class 1. Thus, based on different machining characteristics, such as tool material and TW, the procedure of splitting of nodes continues until the terminal nodes are reached and the nodes can no further be split. It can be clearly noticed that all the terminal nodes are as pure as possible, containing no misclassification error. It is also observed that all the six USM parameters, i.e. CC, TW, TP, tool material, GS and PR are responsible for performing the split operation. Hence, these six parameters influence the MRR response, out of which PR has the major contribution on MRR.

From the developed DT, the following decision rules can be framed to aid the classification process:

Rule 1: If PR $\leq$ 50%, then MRR $\leq$ 0.0150 g/min.

Rule 2: If PR > 50% and GS $\leq$ 410 (mesh), then MRR > 0.0150 g/min.

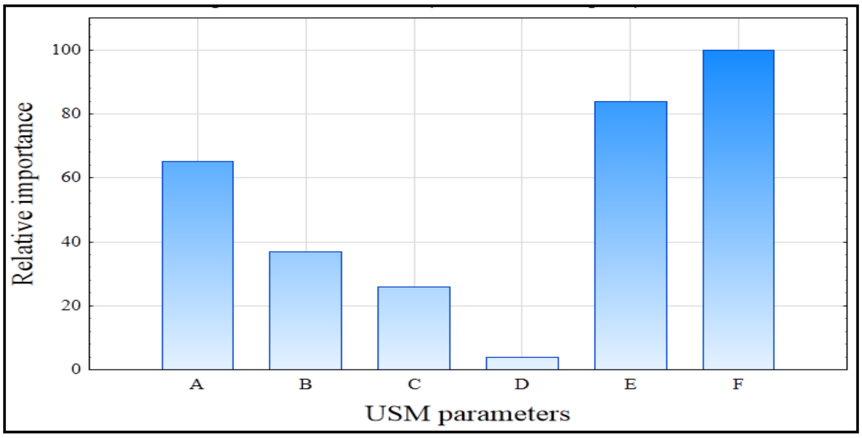
Rule 3: If PR > 50% and GS > 410 (mesh), then MRR $\leq$ 0.0150 g/min.

Rule 4: If PR > 50%, GS $\leq$ 410 (mesh) and TP = solid, then MRR > 0.0150 g/min.

Rule 5: If PR > 50%, GS $\leq$ 410 (mesh), TP = solid and tool material = stainless steel, then MRR $\leq$ 0.0150 g/min.

Rule 6: If PR > 50%, GS $\leq$ 410 (mesh), TP = solid, tool material = stainless steel and TW $\leq$ 4 mm, then MRR > 0.0150 g/min.

Rule 7: If PR > 50%, GS $\leq$ 410 (mesh), TP = solid, tool material = stainless steel, TW > 4 mm and tool material = silver steel, then MRR > 0.0150 g/min.



A-Tool profile, B-Tool material, C-Cobalt content, D-Thickness of work, E- Grit size, F-Power rating

Figure 3.23 Importance of various USM parameters on MRR

It is interesting to note that when the univariate split is performed, all the USM parameters are rated on a 0-100 scale according to their potential importance on the responses. In this case, it can be revealed from Figure 3.23 that PR plays a vital role in controlling the response MRR. On the other hand, CC and TW are the least significant attributes in the said classification process.

Similarly, for TWR, the corresponding DT is developed, which is exhibited in Figure 3.24, to determine the effects of six USM parameters on TWR. In this figure, the datasets with TWR less than or equal to 0.0037 g/min are termed as ‘Class 1’ label, whereas, the datasets with TWR values greater than 0.0037 g/min are grouped as ‘Class 2’ label (where 0.0037 is the calculated average value of TWR). Initially, at the root node, all the instances are primarily classified into Class 1 label, as there are more number of Class 1 values. In the first split, datasets with PR values of less than or equal to 66.049% are subjected to node number 2 and labelled as Class1. There are 24 such datasets, as shown in node 2. On the other hand, datasets with PR values greater than 66.049% are grouped as Class 2 label at node number 3. With GS (mesh) as the splitting criterion, node 3 is further split in which 8 datasets with GS less than or equal to 405.95, classified as Class 2, are sent to node 6, and 4 items having GS greater than 405.95 are merged into node number 7 and labelled as Class 1. At this point, node number 6 represents a terminal node. In this pattern, based on other splitting variables, such as TP, tool material and CC, the process of classification continues until all the terminal nodes arrive where no further splitting is possible. In the developed tree, there are 11 splits and 12 terminal nodes.

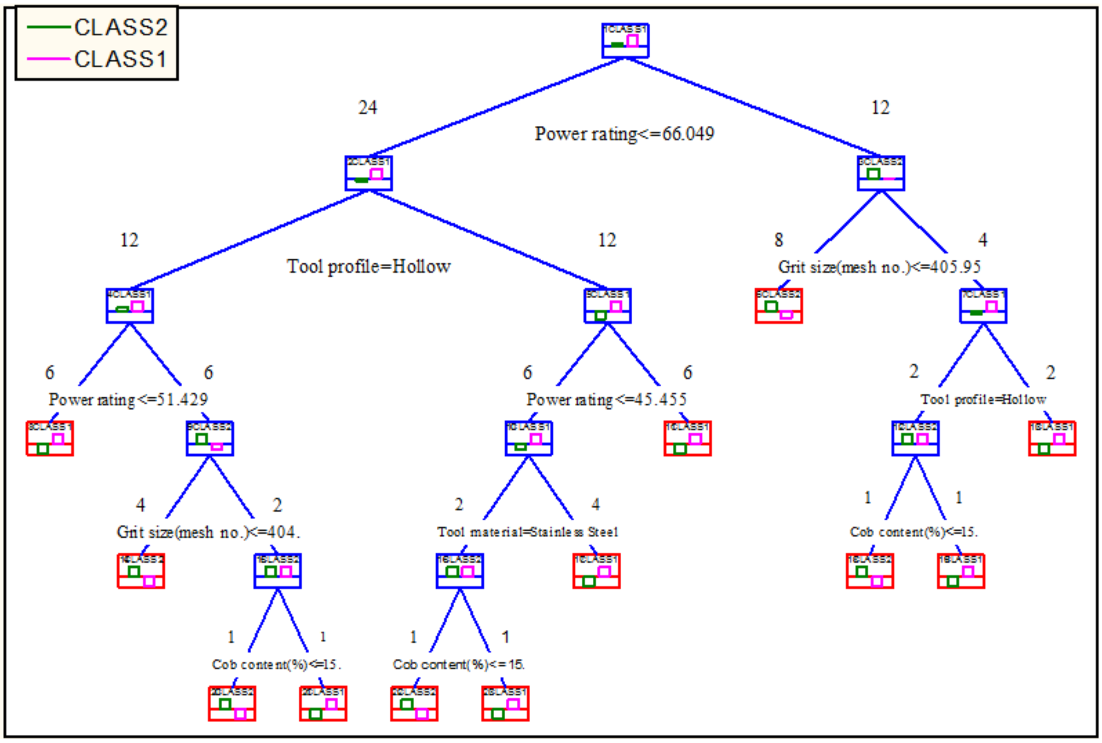


Figure 3.24 CART-based DT for TWR

From the developed tree, the following decision rules can be framed for TWR:

Rule 1: If PR ≤ 66.049%, then TWR ≤ 0.0037 g/min.

Rule 2: If PR > 66.049% and GS ≤ 405.95, then TWR > 0.0037 g/min.

Rule 3: If PR > 66.049%, GS > 405.95 and TP = hollow, then TWR ≤ 0.0037 g/min.

Rule 4: If PR > 66.049%, GS > 405.95, TP = hollow and CC ≤ 15%, then TWR > 0.0037 g/min.

Rule 5: If PR ≤ 66.049%, TP = hollow and tool material = stainless steel, then TWR ≤ 0.0037 g/min.

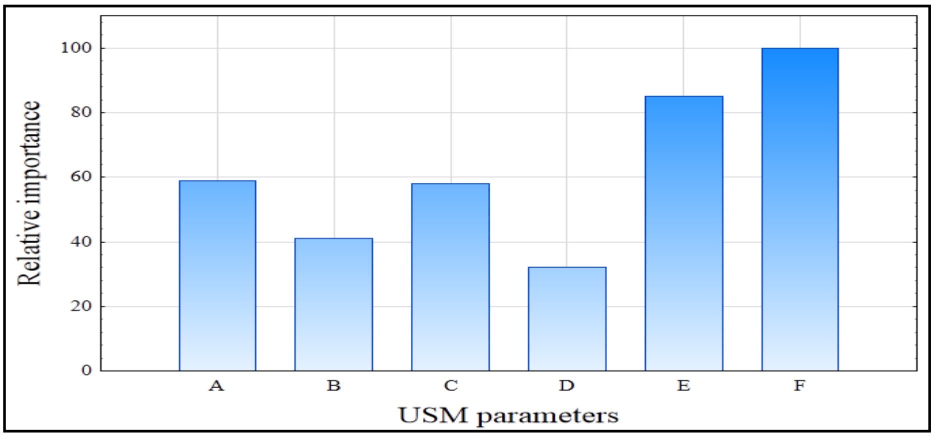
Rule 6: If PR ≤ 66.049%, TP = hollow, tool material = stainless steel and CC ≤ 15 %, then TWR ≤ 0.0037 g/min.

Rule 7: If PR ≤ 66.049%, TP = hollow, PR ≤ 51.429, then TWR ≤ 0.0037 g/min.

Rule 8: If PR ≤ 66.049%, TP = hollow, PR > 51.429 and GS ≤ 404, then TWR > 0.0037 g/min.

Rule 9: If PR ≤ 66.049%, TP = hollow, PR > 51.429, GS > 404 and CC ≤ 15%, then TWR > 0.0037 g/min.

Rule 10: If PR ≤ 66.049%, TP = hollow, PR > 51.429, GS > 404 and CC > 15%, then TWR ≤ 0.0037 g/min.



A-Tool profile, B-Tool material, C-Cobalt content, D-Thickness of work, E- Grit size, F-Power rating

Figure 3.25 Importance of various USM parameters on TWR

Figure 3.25 shows the importance of different USM parameters in influencing TWR, and it becomes quite apparent that PR has the maximum contribution in controlling TWR, followed by GS, TP, CC and tool material.

Thus, in this chapter it can be observed that these DM tools can be effectively applied to all the traditional and non-traditional machining processes to investigate the contributions of varied input parameters on the responses and identify the most suitable parametric combinations for exploring their fullest machining potential. But, the developed decision trees are highly unstable compared to other decision predictors as a small variation in data may result in a major change in the structure of the decision trees, conveying different pictures from the expected ones.

PREDICTIVE MODELING OF AN EDM PROCESS

4. PREDICTIVE MODELING OF AN EDM PROCESS

The main aim of this research work also focuses on applications of four predictive modeling techniques, i.e. XGB, KRG, RBF and GEP to envisage responses of an EDM process. Additionally, it aims to compare the prediction performance of those techniques using four statistical metrics, i.e. MSE, RMSE, R^2 and R^2_{adj}. To achieve this, the experimental dataset provided by Soundhar et al. [128] is considered here. The experiments were designed based on RSM of face centred central composite design plan. Soundhar et al. [128] performed 30 experiments using a die-sinking EDM setup to investigate influences of four factors, such as voltage (V) (in V), current (I) (in A),T_{on} (in µs), and T_{off} (in µs), on MRR, electrode wear rate (EWR) and SR. This investigation considered titanium alloy Ti-13Zr-13Nb (TZN) as the work material. All the EDM parameters during the experiments were set at three different levels, i.e. V 50-60-70 V, I 8-12-16 A, T_{on} 6-8-10 µs, and T_{off} 7-9-11 µs. Table 4.1 represents the detailed experimental design plan along with the measured responses. Among the four responses, MRR is the only LTB type of quality characteristic, while the remaining two are of STB type. In this work, Python 3.7.12 is employed for developing the relevant codes for the considered prediction models. Data handling is performed using Panda (ver.1.1.5) and Numpy (ver. 1.19.5). For model training, XGB (ver. 0.90) and Scikit-learn library (ver. 1.0.2) are utilized.

The applications of predictive modeling techniques and DM algorithms begin with a training dataset. The efficiency and effectiveness of a predictive model is largely dependent on the quality of the training data. Consequently, the training data must encompass all the features of the design space being analyzed. It is essential to develop a training database that represents the design space without any bias. Several sampling techniques can be employed to achieve this. Here, random sampling technique is used since it has several advantages, such as minimal chance of error, simplest way of data generation, equal chance of selection of data points, minimum biasness, less knowledge requirement etc. for training and testing purposes. Training data points are used to build and develop a model, while testing data points are employed to evaluate the model's prediction accuracy on a new set of inputs. In this case, as detailed in Table 4.1, 80% of the observations are randomly selected for training the models, while the remaining 20% are taken as testing data points. These testing points are utilized to assess the model's generalization and performance. It is important to note that the same training and testing data sets are considered here for comparing the prediction performance of different models.

For each of the modeling techniques adopted here, selection of the most appropriate set of hyper-parameters during the training process plays a vital role for the ML or DM models. These hyper-parameters of the models can be selected by trial and error or using suitable optimization methods. Proper combination of these hyper-parameters results in development

Table 4.1 Experimental dataset [128]

Exp. No.	Current (I)	Voltage (V)	Pulse-on time (μs)	Pulse-off time (μs)	MRR (g/min)	SR (μm)	EWR (g/min)
1	8	70	10	7	0.230	11.5580	0.0030
2	12	60	8	9	0.473	14.7170	0.0070
3	12	60	8	9	0.441	14.8670	0.0050
4	8	70	6	7	0.0789	7.6470	0.0040
5	8	50	6	11	0.5075	6.2450	0.0004
6	16	50	10	7	0.2574	14.5140	0.0115
7	12	60	10	9	0.6162	13.6080	0.0070
8	16	70	6	11	0.0860	10.1680	0.0040
9	12	60	8	9	0.4731	10.3250	0.0060
10	12	60	8	9	0.4482	15.8510	0.0080
11	8	70	6	11	0.4272	9.0400	0.0004
12	16	70	10	11	0.6160	14.5140	0.0117
13	12	60	8	9	0.4623	16.2400	0.0170
14	16	60	8	9	0.5707	11.7280	0.0101
15	12	60	8	9	0.4572	12.4850	0.0070
16	16	50	6	7	0.2193	10.0080	0.0080
17	12	60	8	7	0.3220	10.0080	0.0080
18	8	50	6	7	0.0990	6.3010	0.0042
19	16	70	6	7	0.3206	9.5770	0.0105
20	12	50	8	9	0.2050	12.6290	0.0076
21	8	50	10	11	1.0510	10.3890	0.0043
22	12	60	8	11	0.6305	12.1960	0.0057
23	12	60	6	9	0.3400	7.5450	0.0044
24	16	50	6	11	0.0448	6.7530	0.0030
25	8	50	10	7	0.2004	13.2890	0.0039
26	8	60	8	9	0.7129	9.1490	0.0041
27	12	70	8	9	0.2412	18.2140	0.0085
28	16	50	10	11	0.5250	16.7580	0.0107
29	16	70	10	7	0.4086	14.8140	0.0133
30	8	70	10	11	1.0208	14.3220	0.0036

of an accurate model with less complexity. Hyper-parameter optimization and cross-validation are the essential steps in developing a robust prediction model. Together they can reduce loss of information and improve overall performance of a model. In this work, while developing the predictive models using XGB, KRG, RBF and GEP techniques, several trials are conducted with different combinations of the hyper-parameters, and those parameters providing the best results are noted. Values of the considered optimized hyper-parameters for each of the modeling techniques are summarized in Table 4.2.

As mentioned earlier, this work involves developing four predictive models, i.e. XGB, KRG, RBF and GEP to forecast MRR, EWR and SR values during an EDM process. Initially

Table 4.2 Values of different optimized hyper-parameters

Modeling technique	Hyper-parameters	Value
XBG	Learning rate	0.05
	N estimator	100
	Maximum depth	6
	α	1
	λ	1
KRG	θ	0.01
	Polynomial	Quadratic
RBF	d_o	3
	Degree of polynomial	2
GEP	Number of chromosomes	15
	Number of genes	7
	Head size	2
	Linking function	Addition
	Mutation rate	0.0025
	Constant per gene	4
	Lower bound	-10
	Upper bound	10

based on the training dataset, an ensemble learning method, in the form of XGB, is employed for envisaging values of MRR, EWR and SR during EDM operation of TZN alloy. To predict responses of the said EDM process using XGB algorithm, the corresponding Python libraries, like Numpy, Pandas sci-kit-learn and matplot lib etc. are utilized for data handling, data pre-processing, model training and data visualization. In order to fit the data into an XGB model, appropriate parameter values need to be selected. These parameters are generally classified into two types, i.e. booster parameters and learning task parameters [129]. Optimal selection of the hyper-parameters in a model is important to avoid both overfitting and underfitting of data. A possible set of values is tested and the model is run on all these values, followed by subsequent evaluation of the statistical metrics to check predictive accuracy of the considered model. The DTs obtained using XGB algorithm for MRR, EWR and SR are depicted in Figures 4.1-4.3 respectively. The obtained DTs for the corresponding responses are more complex due to the higher value of maximum depth as 6. On the other hand, Tables 4.3-4.5 represent values of MRR, EWR and SR as predicted by the developed XGB algorithm, respectively.

Similarly, using suitable values of the tuning parameters, the corresponding KRG and RBF models are also developed to predict the specified responses. For the KRG model, the number of optimization runs is determined using the elbow method. It is observed that conducting seven optimization runs would yield the most satisfactory results, providing a higher R^2 value with lower computational cost. During modeling using RBF, 15 trials are performed with the basic function scaling parameter (d_o) varying between 1 and 10, and

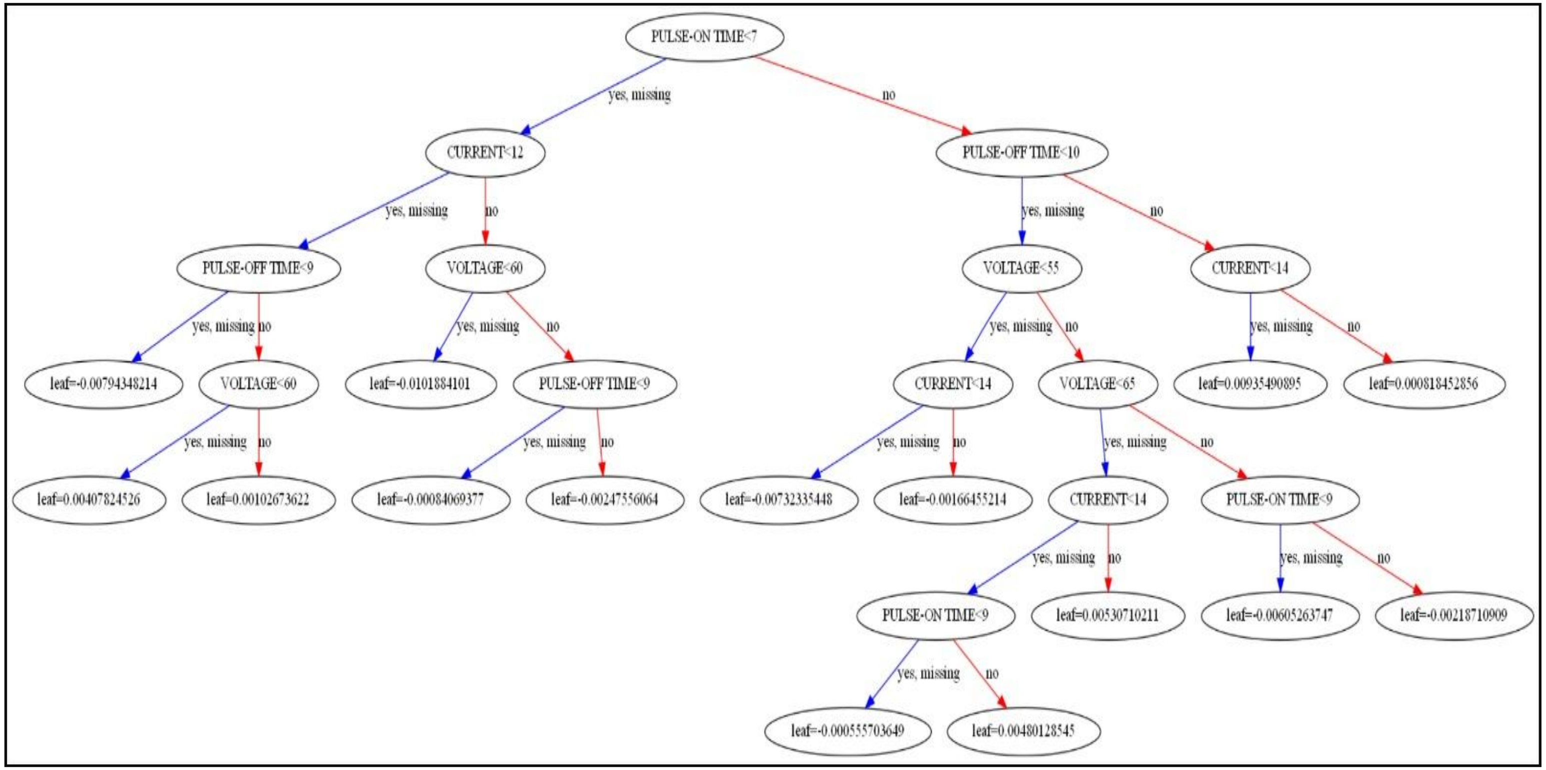


Figure 4.1 Sample tree for MRR

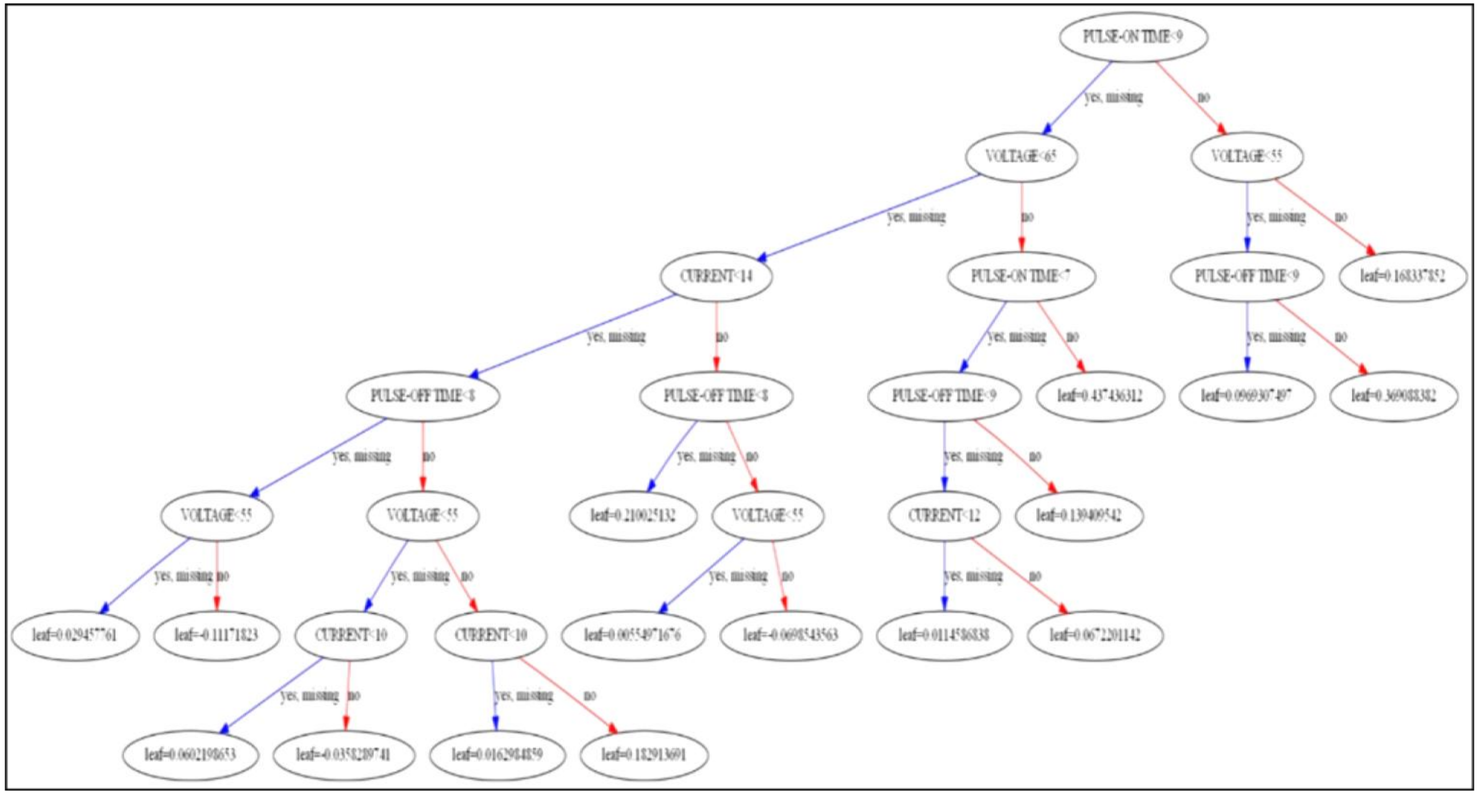


Figure 4.2 Sample tree for EWR

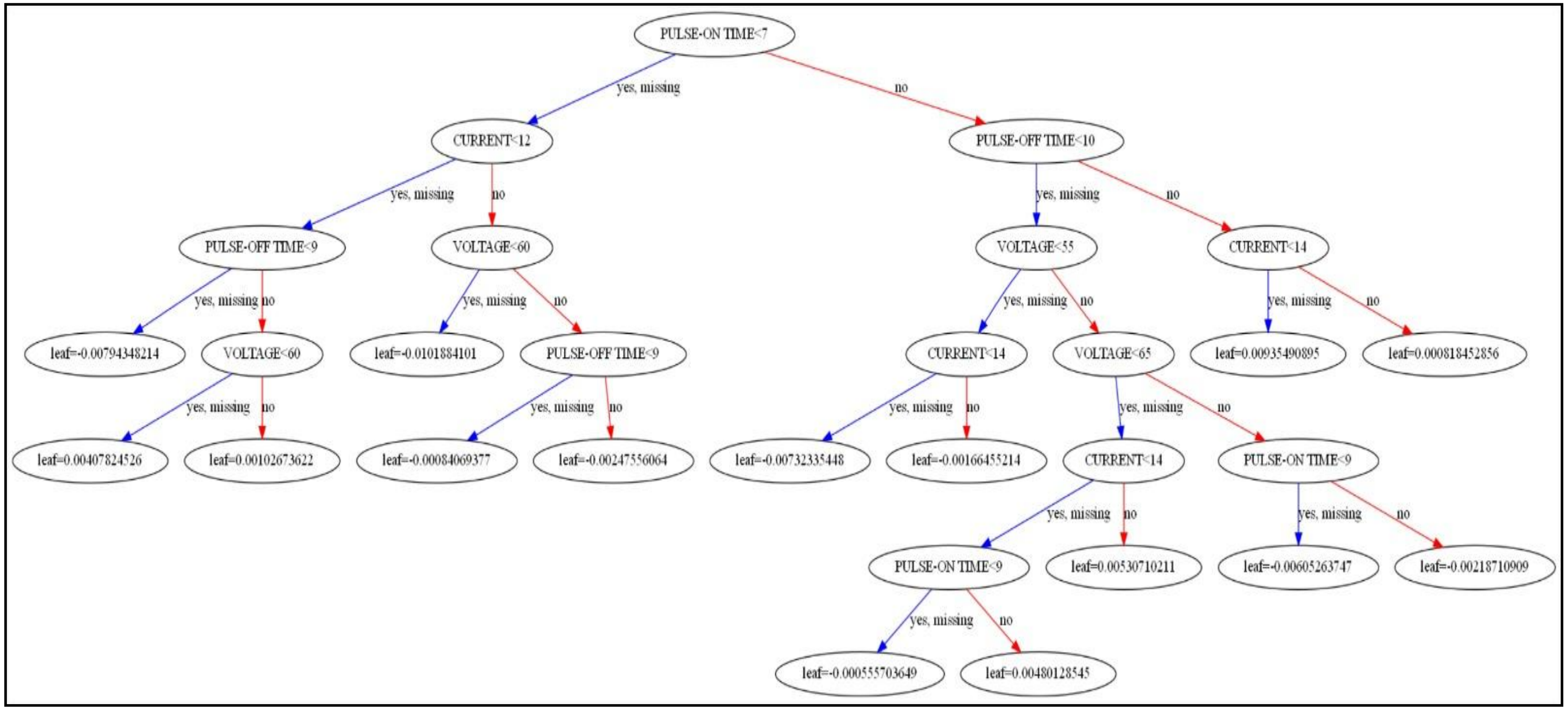


Figure 4.3 Sample tree for SR

polynomial degrees of 1 and 2. The best RBF performance, in terms of maximum R^2 value, is obtained with $d_0 = 6$ and a polynomial degree of 2.

On the other hand, for constructing GEP-based models, a software known as Gen Xpro Tools 5.0 is employed. The population size (chromosomes) defines the run duration of the program. The architecture of a GEP model is determined by the number of genes and size of the head. The size of the head influences complexity of each expression tree by incorporating the sum of sub-expression trees within the model. It employs four basic arithmetical operators, i.e. −, ÷, + and ×. The linking function utilized in GEP is addition, along with other arithmetic operations, to connect the mathematical terms encoded in each gene. During development of these GEP models, all the four genetic operators, i.e. mutation, recombination, inversion and transposition are employed. Figures 4.4-4.6 illustrate the expression trees for MRR, EWR and SR, respectively.

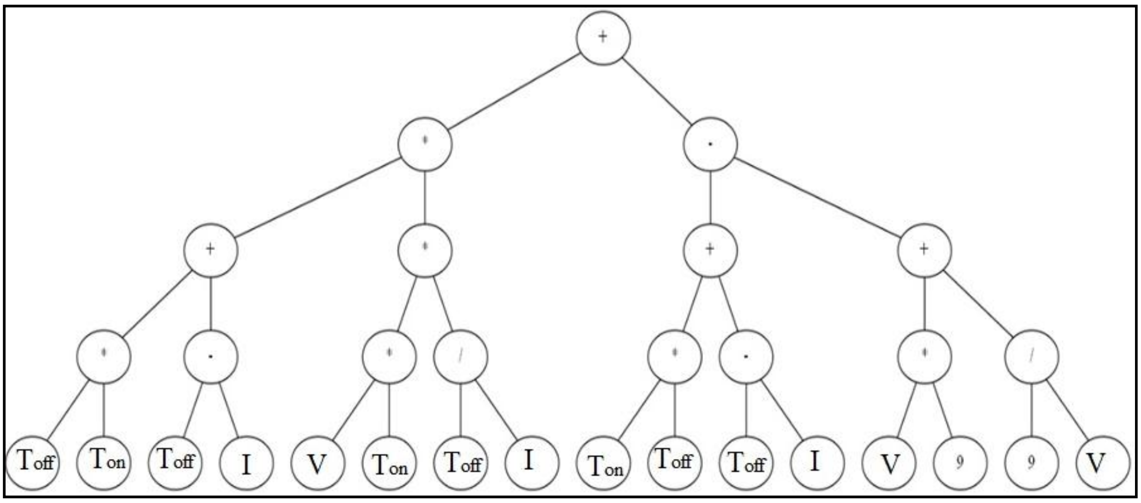


Figure 4.4 Expression tree for MRR

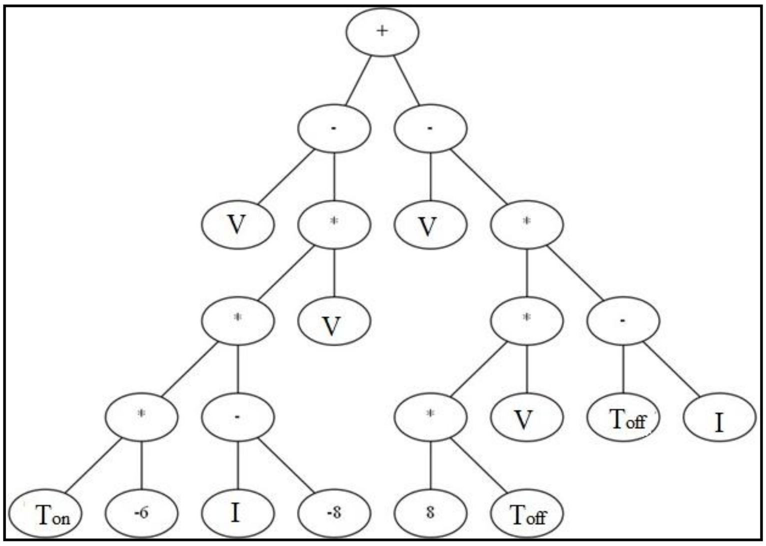


Figure 4.5 Expression tree for EWR

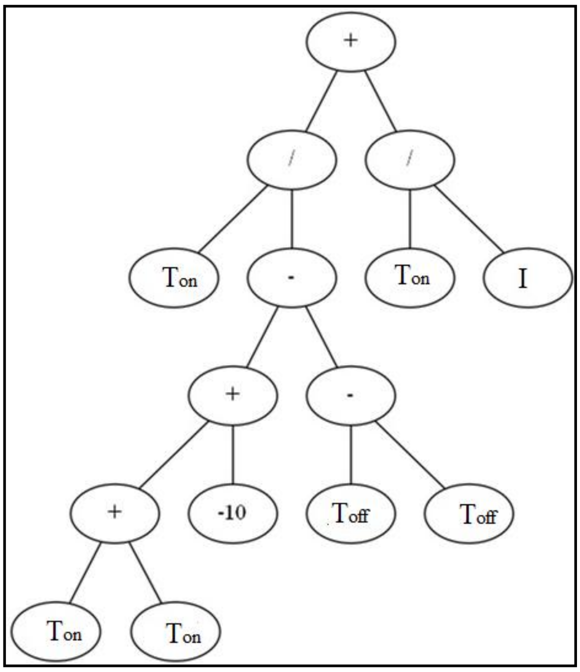


Figure 4.6 Expression tree for SR

These expression trees facilitate formation of the GEP-based models for MRR, EWR and SR, as shown in Eqs. (4.1), (4.2) and (4.3), respectively.

$$MRR = 0.1836 + 7.9944 \times 10^{-6} \times \left(\frac{I \times V\left(-I + T_{off} \times T_{on} + T_{off} - 9 \times V\right) - 9 \times I + T_{off} \times T_{on} \times V \times 2\left(-I + T_{off} \times T_{on} + T_{off}\right)}{I \times V} \right) \tag{4.1}$$

$$EWR = 2.214 \times 10^{-7} \times V \times \left(4 \times T_{off}\left(1 - T_{off}\right) + 3 \times T_{on} \times \left(I + 8\right) + 1\right) - 0.000987 \tag{4.2}$$

$$SR = 20.9702 - \frac{1.1810 \times T_{on} \times \left(I + 2 \times T_{on} - 10\right)}{I \times \left(T_{on} - 5\right)} \tag{4.3}$$

After developing these models, the predicted results are compared with the actual data. Tables 4.3-4.5 exhibit the actual and predicted values of MRR, EWR and SR, respectively during EDM process for all the considered modeling techniques. To analyze the corresponding prediction performance, the actual and predicted values for MRR, EWR and SR are plotted in Figures 4.7-4.9, respectively to visualize closeness of the data points. Stronger predictive ability is indicated when the experimental and predicted results are close to each other. It can be observed from these figures that all the developed models are quite potential to predict the considered responses within a ± 15% error range for the EDM process. Figure 4.7 shows that for MRR prediction, all the models provide nearly ideal estimates, with most of the data points closely aligning with the diagonal line. However, in case of GEP model, few data points are beyond the ± 15% error range.

Table 4.3 Actual and predicted values of MRR for the developed models

Exp. No.	Actual	XGB	KRG	RBF	GEP
1	0.23	0.2304	0.2310	0.2300	0.5170
2	0.473	0.4594	0.4637	0.4612	0.3784
3	0.441	0.4594	0.4637	0.4612	0.3784
4	0.0789	0.0802	0.0793	0.0789	0.2994
5	0.5075	0.5063	0.5074	0.5075	0.4081
6	0.2574	0.2581	0.2581	0.2574	0.2872
7	0.6162	0.6159	0.6089	0.6162	0.4930
8	0.086	0.0865	0.0864	0.0860	0.3199
9	0.4731	0.4594	0.4637	0.4612	0.3784
10	0.4482	0.4594	0.4637	0.4612	0.3784
11	0.4272	0.4269	0.4280	0.4272	0.4977
12	0.616	0.6160	0.6185	0.6160	0.5834
13	0.4623	0.4594	0.4637	0.4612	0.3784
14	0.5707	0.5698	0.5672	0.5707	0.3201
15	0.4572	0.4594	0.4637	0.4612	0.3784
16	0.2193	0.2194	0.2183	0.2193	0.2149
17	0.322	0.3228	0.3138	0.2832	0.2939
18	0.099	0.0991	0.0980	0.0990	0.2664
19	0.3206	0.3199	0.3201	0.2451	0.2273
20	0.205	0.2054	0.1977	0.1813	0.3461
21	1.051	1.0094	1.0525	1.0510	0.8020
22	0.6305	0.6301	0.6232	0.6305	0.4860
23	0.34	0.1043	0.2917	0.2891	0.2898
24	0.0448	0.0465	0.0453	0.0448	0.2811
25	0.2004	0.2057	0.2003	0.2004	0.4219
26	0.7129	0.7125	0.7091	0.7129	0.4950
27	0.2412	0.2416	0.2340	0.2412	0.4108
28	0.525	0.5252	0.5281	0.5164	0.4694
29	0.4086	0.4083	0.4093	0.4086	0.3284
30	1.0208	1.0194	1.0223	0.9514	1.0490

Table 4.4 Actual and predicted values of EWR for the developed models

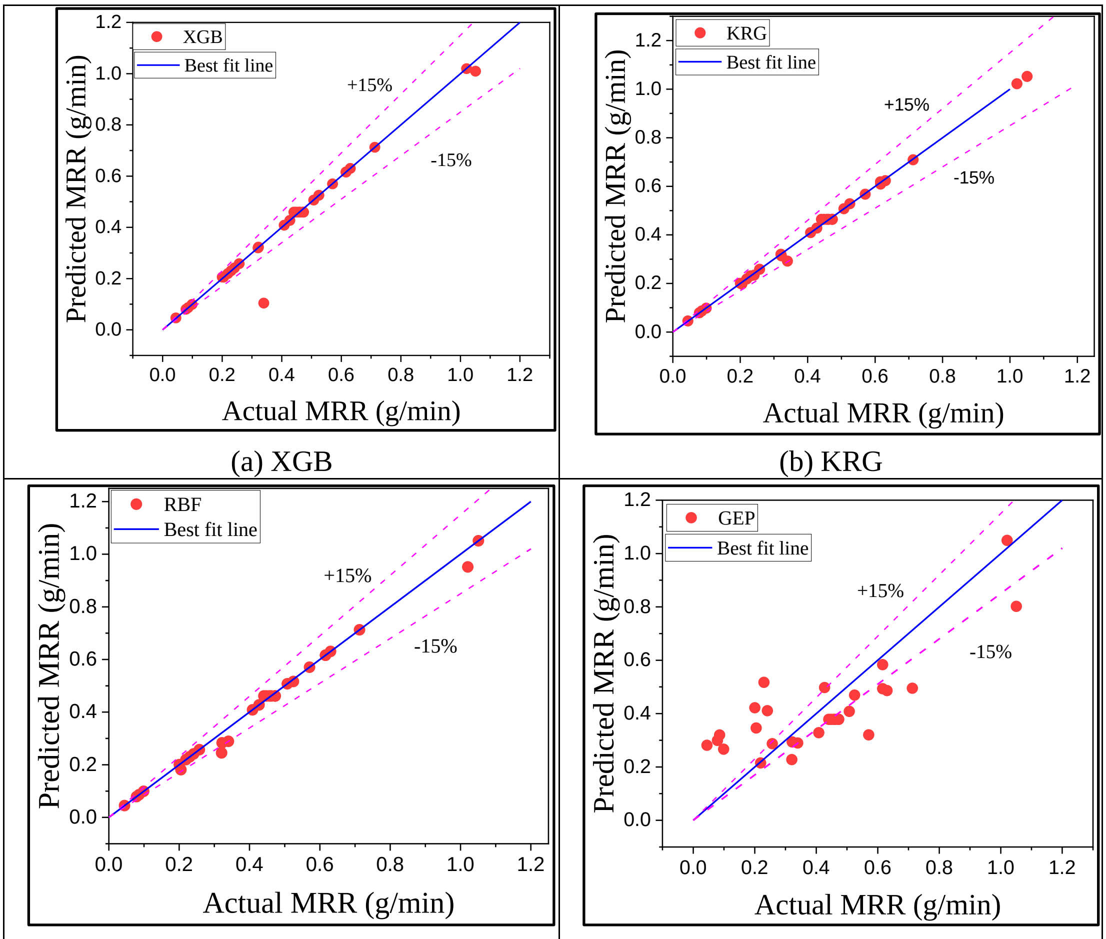
Exp. No.	Actual	XGB	KRG	RBF	GEP
1	0.0030	0.0002	0.0031	0.0038	0.0069
2	0.0070	0.0005	0.0081	0.0080	0.0068
3	0.0050	0.0005	0.0081	0.0080	0.0068
4	0.0040	0.0001	0.0039	0.0036	0.0039
5	0.0004	0.0005	0.0006	0.0010	0.0008
6	0.0115	0.0003	0.0116	0.0114	0.0098
7	0.0070	0.0006	0.0077	0.0092	0.0084
8	0.0040	0.0001	0.0040	0.0051	0.0091
9	0.0060	0.0005	0.0081	0.0080	0.0068
10	0.0080	0.0005	0.0081	0.0080	0.0068
11	0.0004	0.0004	0.0001	0.0006	0.0014
12	0.0117	0.0006	0.0114	0.0109	0.0136
13	0.0170	0.0005	0.0081	0.0080	0.0068
14	0.0101	0.0006	0.0104	0.0100	0.0100
15	0.0070	0.0005	0.0081	0.0080	0.0068
16	0.0080	0.0002	0.0081	0.0081	0.0066

Table 4.4 Contd.

Exp. No.	Actual	XGB	KRG	RBF	GEP
17	0.0080	0.0003	0.0089	0.0092	0.0073
18	0.0042	0.0001	0.0042	0.0041	0.0025
19	0.0105	0.0003	0.0097	0.0100	0.0096
20	0.0076	0.0002	0.0087	0.0070	0.0055
21	0.0043	0.0010	0.0041	0.0035	0.0029
22	0.0057	0.0006	0.0063	0.0064	0.0060
23	0.0044	0.0001	0.0043	0.0063	0.0052
24	0.0030	0.0000	0.0026	0.0030	0.0062
25	0.0039	0.0002	0.0036	0.0042	0.0046
26	0.0041	0.0007	0.0044	0.0041	0.0037
27	0.0085	0.0002	0.0091	0.0078	0.0081
28	0.0107	0.0005	0.0102	0.0088	0.0094
29	0.0133	0.0004	0.0130	0.0134	0.0141
30	0.0036	0.0010	0.0033	0.0033	0.0044

Table 4.5 Actual and predicted values of SR for the developed models

Exp. No.	Actual	XGB	KRG	RBF	GEP
1	11.5580	11.5580	11.9127	11.5580	12.8244
2	14.7170	14.2240	13.2482	14.1387	13.7295
3	14.8670	14.2240	13.2482	14.1387	13.7295
4	7.6470	7.6471	8.4547	7.6470	7.3938
5	6.2450	6.2447	5.4725	6.2450	7.3938
6	14.5140	14.5138	15.4780	14.5140	15.0871
7	13.6080	13.6102	14.2643	13.6080	14.3329
8	10.1680	10.1664	10.1398	10.1680	8.7515
9	10.3250	14.2240	13.2482	14.1387	13.7295
10	15.8510	14.2240	13.2482	14.1387	13.7295
11	9.0400	9.0401	9.6170	9.0400	7.3938
12	14.5140	14.5142	16.1467	14.5140	15.0871
13	16.2400	14.2240	13.2482	14.1387	13.7295
14	11.7280	11.7304	12.3706	11.7280	14.3329
15	12.4850	14.2240	13.2482	14.1387	13.7295
16	10.0080	10.0033	9.1178	10.0080	8.7515
17	10.0080	10.0123	11.5877	11.6635	13.7295
18	6.3010	6.3012	6.2871	6.3010	7.3938
19	9.5770	9.5778	9.8236	11.4478	8.7515
20	12.6290	12.6308	15.4474	10.2441	13.7295
21	10.3890	11.5693	11.7622	10.3890	12.8244
22	12.1960	12.1980	12.0106	12.1960	13.7295
23	7.5450	10.0006	8.7093	8.9963	8.2989
24	6.7530	6.7538	7.4834	6.7530	8.7515
25	13.2890	10.2138	11.5477	13.2890	12.8244
26	9.1490	9.1510	9.9208	9.1490	12.5227
27	18.2140	18.2154	17.1945	18.2140	13.7295
28	16.7580	16.7573	15.3191	13.4835	15.0871
29	14.8140	14.8134	14.5651	14.8140	15.0871
30	14.3220	14.3214	14.1303	13.3879	12.8244



(a) XGB (b) KRG

(c) RBF (d) GEP

Figure 4.7 Comparison between the actual and predicted values for MRR by XGB, KRG, RBF and GEP

The performance of XGB, KRG and RBF is noticed to be similar with good prediction. In case of EWR prediction by these models (Figure 4.8), a trend similar to MRR is observed. The XGB and KRG models present quite similar prediction behaviour, whereas, most of the poor predictions are noticed for GEP model. Further, it can be noticed from Figure 4.9 that for SR prediction, XGB and RBF show good prediction results as compared to GEP, whereas, it is interesting to note that KRG provides quite acceptable prediction values than those of GEP.

Figure 4.10 shows the comparisons between the actual and predicted values for MRR, EWR and SR across all the four prediction tools. This figure reveals that the predicted values for MRR, EWR and SR from the developed models closely align with their actual values. On the other hand, the prediction performance of GEP in case of SR and XGB in case of EWR are poor. After critically analyzing Figure 4.10, it can be unveiled that while predicting the considered response values, the relative performance of XGB, KRG and RBF are quite better as compared to GEP for the considered EDM operation.

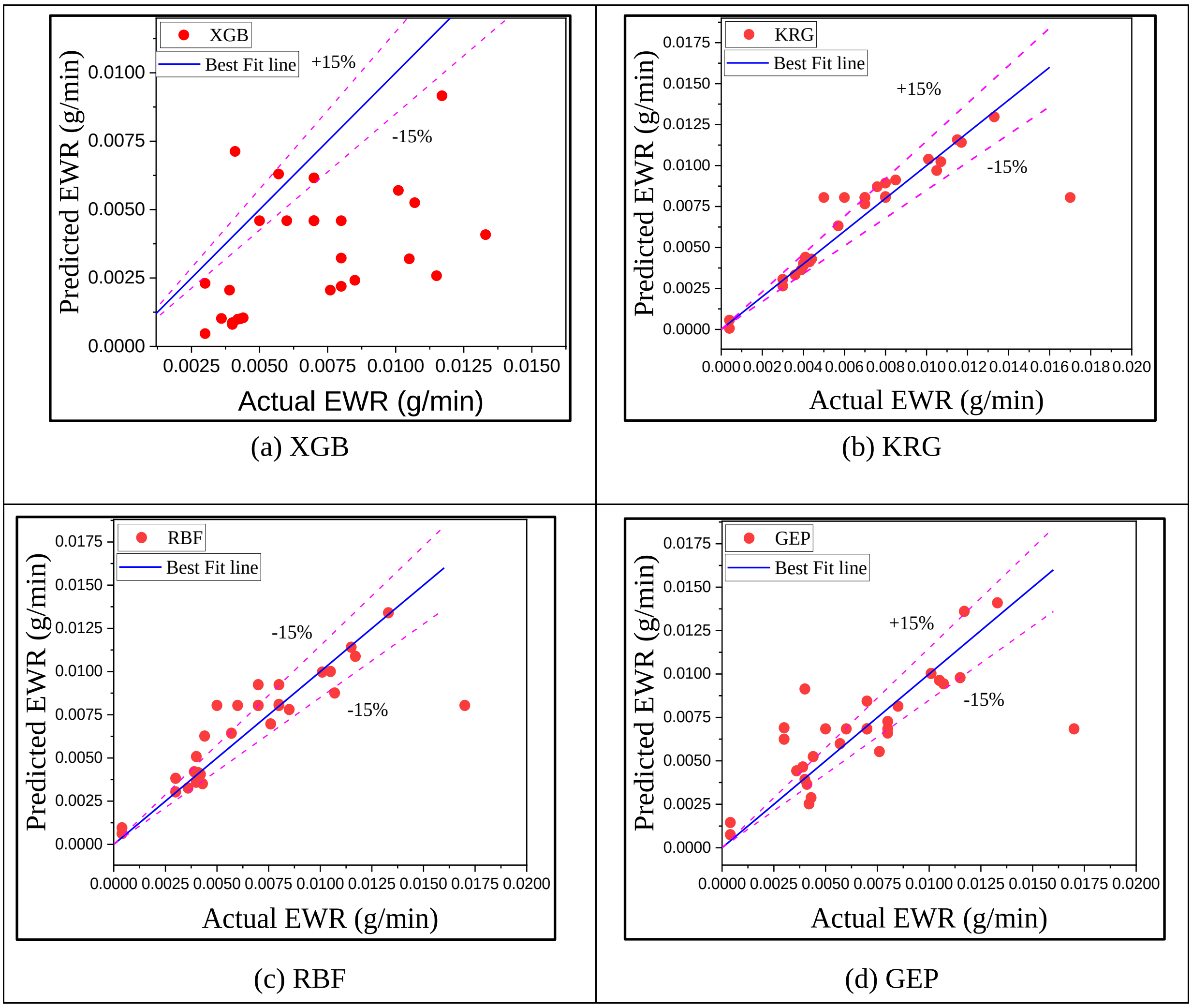


Figure 4.8 Comparison between the actual and predicted values for EWR by XGB, KRG, RBF and GEP

For all the four models, the predicted values for MRR, EWR and SR have considerably low deviations from the corresponding experimental values. The predicted values of MRR can be seen closer to the actual values for XGB, KRG and RBF. On the other hand, GEP provides average prediction results with moderate deviations of the MRR values. Similar observations can be noticed for EWR. However, application of XGB shows lower prediction performance for EWR, but it is efficient for MRR and SR. On the other hand, XGB, RBF and KRG models show better prediction results for SR as compared to GEP. The overall application of GEP shows average prediction results for all the considered responses of the said EDM process.

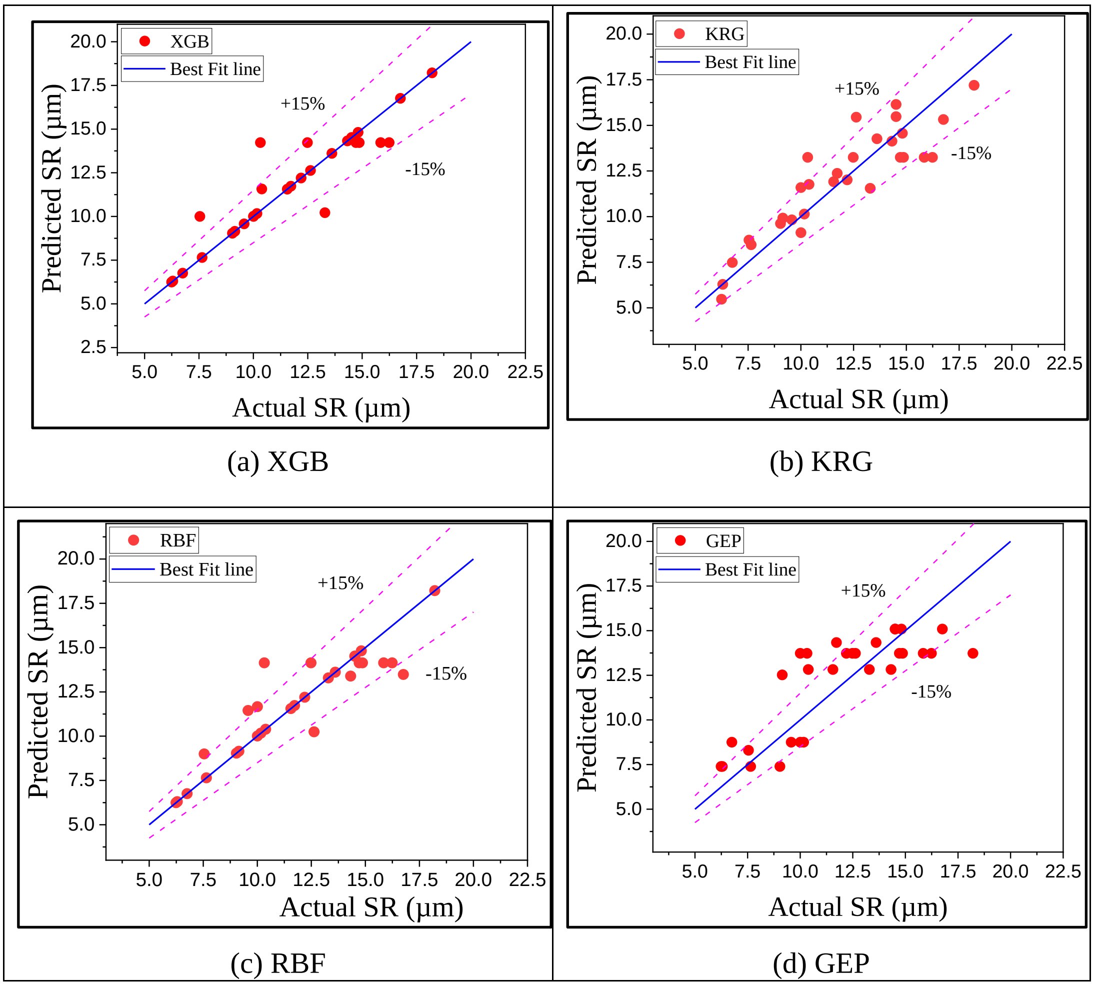


Figure 4.9 Comparison between the actual and predicted values for SR by XGB, KRG, RBF and GEP

However, it would be quite interesting to decide which of the developed models performs best for the considered EDM process. Hence, to overcome this issue, their performance is evaluated using four statistical metrics, i.e. MSE, RMSE, R^2 and R^2_{adj}. To determine whether a model is overfitting or underfitting, its performance is evaluated on both the training and test data. In order to validate prediction accuracy of all the four modeling techniques for this EDM process, the corresponding statistical metrics are calculated. Table 4.6 provides values of MSE, RMSE, R^2 and R^2_{adj}, when XGB, KRG, RBF and GEP models are employed for prediction of MRR, EWR and SR values during the EDM operation. These comparison results for MRR, EWR and SR values are also pictorially presented in Figures 4.11-4.13, respectively. Among these metrics, lower values of MSE and RMSE are always preferred, while higher values of R^2 and R^2_{adj} are recommended for assessing performance of the prediction tools. The excellent results across all these statistical metrics strongly demonstrate effectiveness and potential of the developed models envisaging response values of the said EDM process.

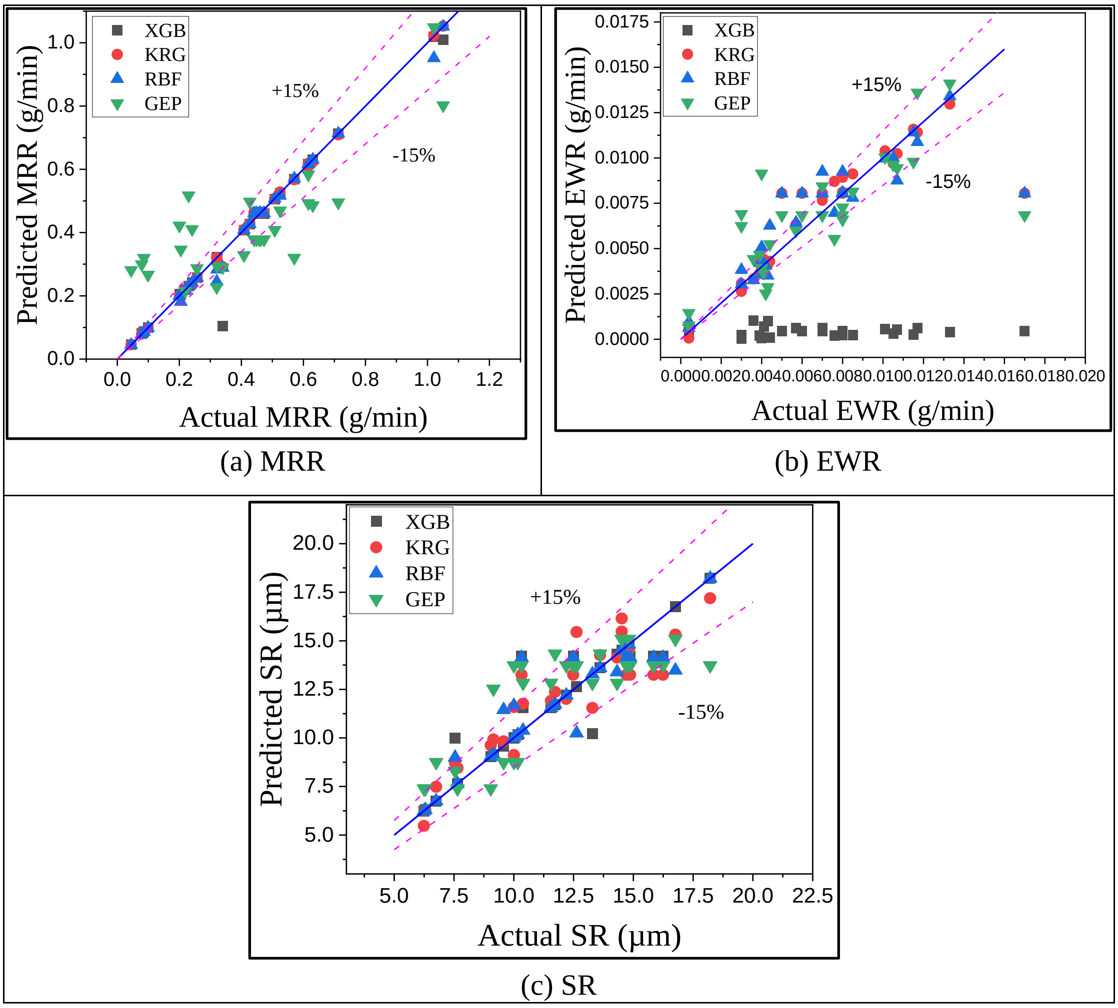


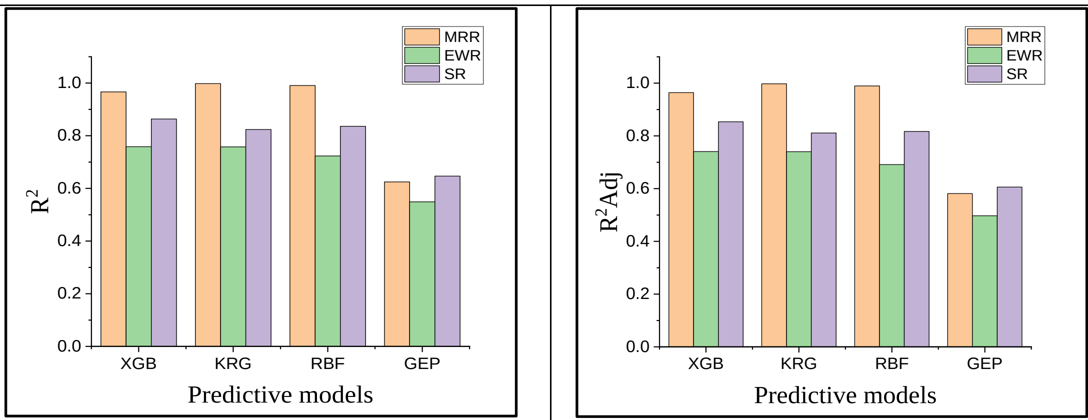
Figure 4.10 Actual and predicted values for MRR, EWR and SR using various models

From these figures, it is evident that XGB, KRG and RBF models outperform GEP in terms of all the statistical measures for the three EDM responses being considered. From Table 4.6, it can be noticed that in case of MRR, the predictive models have the best values of R^2 and R^2_{adj} as 0.9666 and 0.9642 for XGB; 0.9979 and 0.9977 for KRG; and 0.9907 and 0.9897 for RBF, respectively, taking into account the overall dataset. The corresponding values of MSE and RMSE as 0.0019 and 0.044 (for XGB), 0.00012 and 0.0110 (for KRG), and 0.00054 and 0.0232 (for RBF) also confirm superior performance of XGB, KRG and RBF models based on the overall dataset. The obtained values of R^2 and R^2_{adj} as 0.6247 and 0.5814, and MSE and RMSE as 0.0218 and 0.1477 represent the poorer prediction performance of GEP. Thus, the prediction of MRR based on the overall dataset reveals that KRG is the best model for prediction, followed by XGB and RBF.

While predicting EWR values, similar results can also be noticed. The XGB, KRG and RBF models show good accuracy with maximum R^2 and R^2_{adj}, and minimum MSE and RMSE values. The KRG emerges out as the most accurate model with maximum R^2 (0.7587) and R^2_{adj} (0.7401), and minimum MSE (3.36E-06) and RMSE (0.0018) values. On the other hand, GEP exhibits the least R^2 (0.5492) and R^2_{adj} (0.4972) values for EWR as compared to other models. Furthermore, from the obtained values of R^2 (0.6467), R^2_{adj} (0.6059), MSE (3.6687) and RMSE (1.9154) while predicting SR, it can be clearly noticed that GEP exhibits the worst prediction performance since its accuracy deteriorates with the testing and overall datasets.

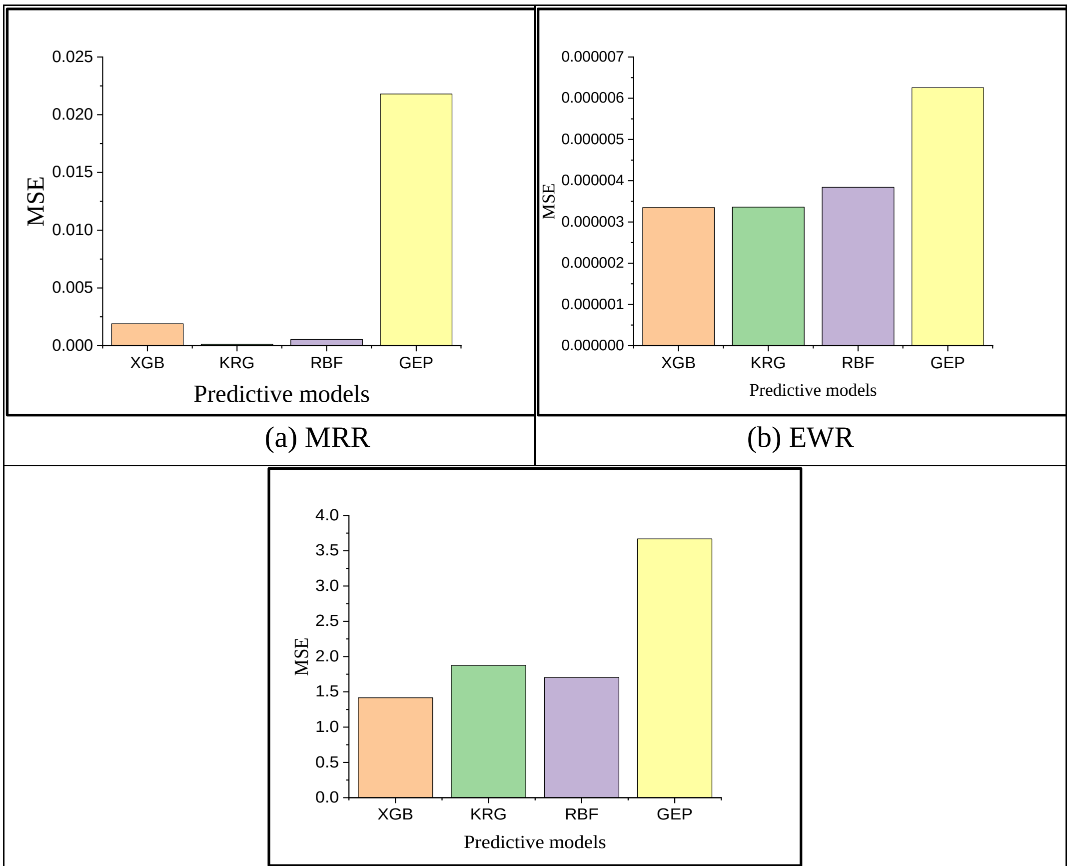
Table 4.6 Values of different statistical metrics for MRR, EWR and SR

Model	Response	Dataset	MSE	RMSE	R^2	R^2_{adj}
XGB	MRR	Testing	0.0095	0.0977	0.8952	0.8875
		Training	3.5767	0.0059	0.9992	0.9992
		Overall	0.0019	0.044	0.9666	0.9642
	EWR	Testing	1.04E-06	0.00101	0.8602	0.8500
		Training	3.93E-06	0.0019	0.7358	0.7166
		Overall	3.35E-06	0.0018	0.7583	0.7407
	SR	Testing	3.0653	1.7508	0.3339	0.2855
		Training	1.0034	1.0017	0.9122	0.9058
		Overall	1.4158	1.1898	0.8636	0.8537
KRG	MRR	Testing	0.0004	0.02042	0.9609	0.9581
		Training	4.73E-05	0.0068	0.9993	0.9993
		Overall	0.00012	0.0110	0.9979	0.9977
	EWR	Testing	6.79E-07	0.0008	0.8528	0.8420
		Training	4.03E-06	0.0020	0.7425	0.7238
		Overall	3.36E-06	0.0018	0.7587	0.7401
	SR	Testing	2.3756	1.5413	0.7255	0.7056
		Training	1.7430	1.3202	0.8437	0.8323
		Overall	1.8748	1.3692	0.8237	0.8109
RBF	MRR	Testing	0.00175	0.0419	0.9638	0.9596
		Training	0.000015	0.0038	0.9997	0.9997
		Overall	0.00054	0.0232	0.9907	0.9897
	EWR	Testing	1.75E-06	0.0013	0.6198	0.5765
		Training	4.36E-06	0.0021	0.7213	0.6892
		Overall	3.84E-06	0.0019	0.7231	0.6912
	SR	Testing	3.2476	1.8021	0.5899	0.5426
		Training	1.0424	1.0210	0.9064	0.8956
		Overall	1.7040	1.3053	0.8359	0.8170
GEP	MRR	Testing	0.0342	0.1850	0.6554	0.6156
		Training	0.0181	0.1344	0.6037	0.5579
		Overall	0.0218	0.1477	0.6247	0.5814
	EWR	Testing	5.4E-06	0.0023	0.2151	0.1245
		Training	6.516E-06	0.0025	0.5929	0.5459
		Overall	6.256E-06	0.0025	0.5492	0.4972
	SR	Testing	4.5016	2.1217	0.4722	0.4113
		Training	3.4153	1.8480	0.6830	0.6464
		Overall	3.6687	1.9154	0.6467	0.6059



(a) R^2 (b) R^2_{adj}

Figure 4.11 Prediction performance of the models in terms of R^2 and R^2_{adj}



(a) MRR (b) EWR

(c) SR

Figure 4.12 Prediction performance of the models with respect to MSE

In case of SR, the predictive models show better values of R^2 and R^2_{adj} as 0.8636 and 0.8537 for XGB; 0.8237 and 0.8109 for KRG; and 0.8359 and 0.8170 for RBF considering the overall dataset. The corresponding values of MSE and RMSE as 1.4158 and 1.1898 (for XGB); 1.8748 and 1.3692 (for KRG); and 1.7040 and 1.3053 (for RBF) represent the

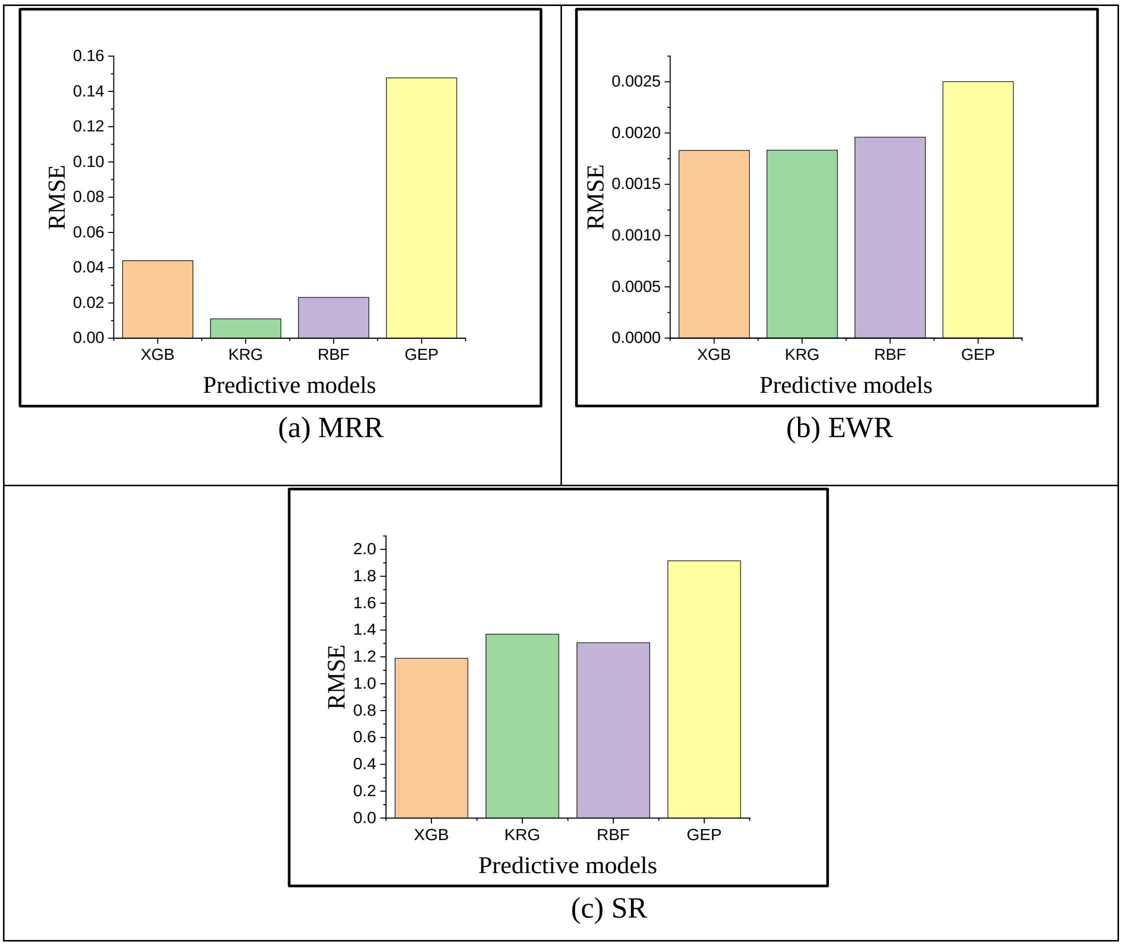


Figure 4.13 Prediction performance of the models with respect to RMSE

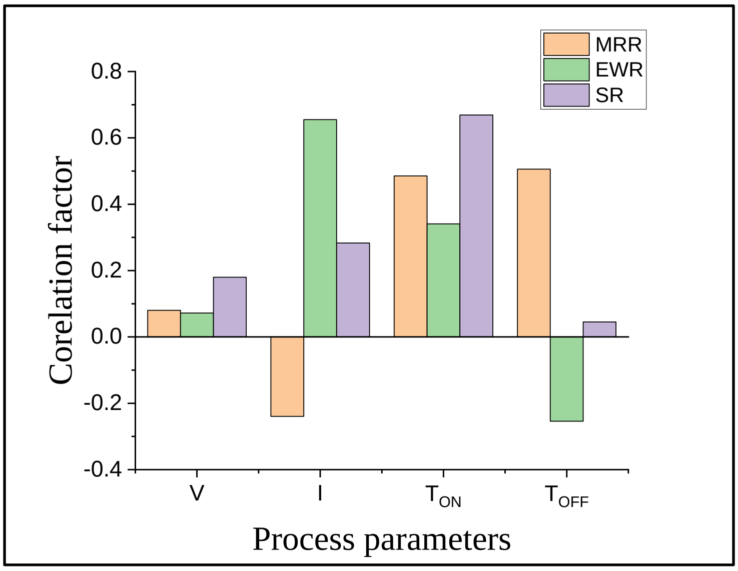


Figure 4.14 Feature importance ranking for MRR, EWR and SR

superior performance of XGB, KRG and RBF models than GEP. However, based on these observations, it can be further stated that GEP shows the worst values of all the four developed metrics for all the responses. Thus, the performance of GEP is found to be inconsistent for MRR, EWR and SR responses. Hence, based on the statistical metrics and overall dataset, it can be concluded that XGB, KRG and RBF outperform GEP in predicting the considered response values for the said EDM process.

In addition, feature importance analysis is also carried out to identify the most influencing EDM parameter affecting the responses. Figure 4.14 displays the results of the feature importance analysis. It reveals that I and T_{on} are the most influencing parameters, followed by V which are in accordance with the previous experimental results. The feature importance analysis is conducted using the optimized hyper-parameters. The importance scores are calculated and aggregated to single out the most influential parameter for the EDM process.

Thus, this chapter reveals that application of predictive techniques in the form of XGB, KRG, RBF and GEP can be effectively employed as the prediction tools for the considered EDM process as well as other varied machining process.

PREDICTION OF EDM PERFORMANCE USING NEURAL NETWORKS

5. PREDICTION OF EDM PERFORMANCE USING NEURAL NETWORKS

Somasundaram and Kumar [130] performed 16 experiments to optimize and study the effects of four input parameters of an EDM process, such as I_P, T_{on}, T_{off} and material of the electrode on SR, MRR, TWR, overcut (OC), taper cut (TC), circularity (CIR) and cylindricity (CYL) during machining of AZ31 alloy. The experiments were designed using Taguchi's L_{16} orthogonal array, and three materials, i.e. copper (Cu), brass (Br) and graphite (Gr) were selected as the electrodes. The size of AZ31 magnesium alloy was considered as $50\times30\times6$ mm. During the experiments, all the four input parameters were set at four different operating values, as depicted in Table 5.1. The detailed experimental design plan with the values of seven responses is exhibited in Table 5.2.

Table 5.1 Different EDM process parameters and their levels [130]

Input parameter	Unit	Level			
		−1	-0.333	0.333	+1
T_{on}	μs	10	20	30	40
T_{off}	μs	5	6	7	8
I_P	A	3	6	9	12
M	-	Cu	Br	Gr	Cu

Table 5.2 Experimental design plan and measured responses for the EDM process [130]

Sl.No.	Coded input parameters				Machining characteristics						
	I_P (A)	T_{on} (μs)	T_{off} (μs)	Electrode material	SR (μm)	MRR (mm^3/min)	TWR (gm/min)	OC (mm)	TC (mm)	CIR (mm)	CYL (mm)
1	-1	-1	-1	-1	3.2	7.5512	0.0042	0.014	0.061	0.0184	0.0421
2	-1	-0.333	-0.333	-0.333	6.68	61.181	0.0206	0.0865	0.045	0.0115	0.0677
3	-1	0.333	0.333	0.333	9.89	89.4419	0.0184	0.1075	0.037	0.008	0.0905
4	-1	1	1	1	11.31	47.7563	0.0014	0.0835	0.037	0.0068	0.0746
5	-0.333	-1	-0.333	0.333	5	34.8201	0.0062	0.0715	0.010	0.041	0.1209
6	-0.333	-0.333	-1	1	4.96	30.4132	0.0015	0.0330	0.059	0.0074	0.0671
7	-0.333	0.333	1	-1	10.02	81.8268	0.0023	0.0625	0.033	0.0735	0.0749
8	-0.333	1	0.333	-0.333	9.79	129.2052	0.0369	0.1	0.039	0.0097	0.0763
9	0.333	-1	0.333	1	6.55	14.9400	0.0010	0.0315	0.002	0.0353	0.0640
10	0.333	-0.333	1	0.333	11.27	73.1567	0.0063	0.1603	0.023	0.0062	0.0814
11	0.333	0.333	-1	-0.333	11.38	131.6526	0.0429	0.155	0.028	0.0456	0.0454
12	0.333	1	-0.333	-1	12.76	156.8343	0.0084	0.214	0.015	0.0023	0.0921
13	1	-1	1	-0.333	4.80	46.3972	0.0255	0.01	0.039	0.0043	0.0452
14	1	-0.333	0.333	-1	8.04	33.2228	0.0010	0.1375	0.006	0.0269	0.0628
15	1	0.333	-0.333	1	9.00	167.2435	0.0091	0.169	0.004	0.0311	0.0780
16	1	1	-1	0.333	14.28	202.6980	0.0116	0.2855	0.004	0.0640	0.1275

Based on the experimental data, four significant NN models, specifically FFNN, CNN, RNN and LSTM are subsequently developed to predict the responses of the said EDM process. Their prediction performance is later compared using four statistical error metrics. The current research work is conducted using the proposed methodologies, which comprises three main stages: (a) data preparation, (b) model development and (c) model validation. In the data preparation phase, the considered experimental dataset is divided into two subsets, i.e. testing and training. 70% of the total data is taken for the training purpose, while the test dataset is made up of the remaining elements. Data preparation improves an NN's performance and reduces its complexity. As a result, data preparation procedures are specifically implemented to accomplish three goals, e.g. (a) identifying and replacing missing values with the mean values, (b) removing duplicate entries and (c) preparing a valid dataset for further processing. During model development stage, the training dataset is used to train the NN model with feed-forward back propagation algorithm. At this stage, the impact of the number of iterations (epochs) and random sampling technique is investigated. In the last stage, the model is validated employing different statistical metrics, such as MSE, RMSE, R^2 and R^2_{adj}. In this research work, the NN models are developed using the Keras library in the tensorflow framework with the help of Python language.

When the data are pre-processed, an NN model performs well. The application of any NN model starts with a set of training data. The Levenberg-Marquardt back-propagation algorithm is used to train the network along with learngdm as the adaptation learning function and tansig as the transfer function. The number of iterations is based on the epochs. Here, 70% of the data is taken for training and 30% for testing purpose while developing various NN models. As mentioned earlier, NNs are made up of interconnected nodes, known as neurons that collaborate to process information, identify patterns and relationships in input data, and make predictions or decisions. The objective of the present study is to develop four NN models, i.e. FFNN, CNN, LSTM and RNN as the prediction tools for an EDM process. Here, inputs to the NN models are T_{on}, T_{off}, I_P and electrode material. The model output parameters are SR, MRR, TWR, OC, TC, CIR and CYL. As the data provided to the NN vary in range, so they are normalized to streamline the training process. Consequently, all the inputs and outputs to the network are transformed. For the EDM operation of AZ31 alloy, the response MRR, being a maximum type of attribute, is optimized for its highest value. Conversely, for attributes, such as SR, TWR, OC, TC, CIR and CYL, their lower values are targeted. For each NN model, the input and output parameters are shown in Figure 5.1.

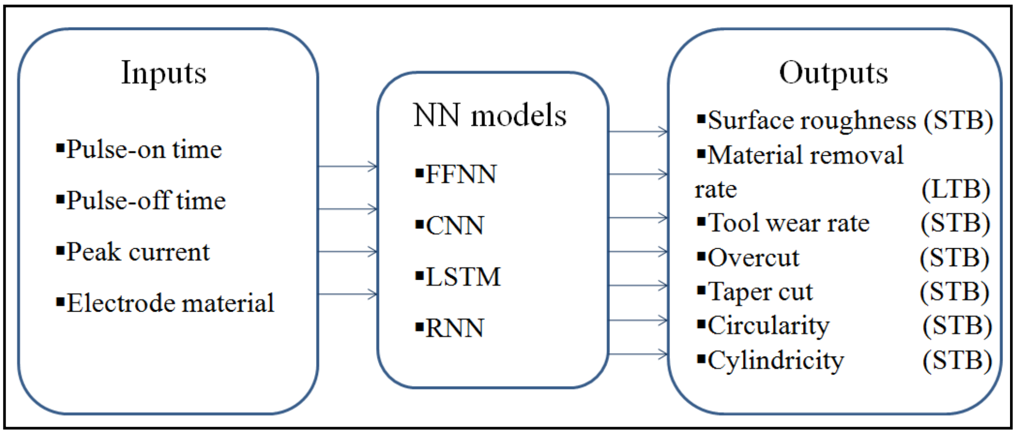


Figure 5.1 Functional representation of the NN

Correspondingly, each NN model's architecture is constructed with different type and number of layers, each containing varying number of nodes. However, each NN model's architecture is developed here taking into account the optimal computational effort.

FFNN is the simplest NN model because the input is always processed in only one direction. In an NN, neurons use an activation function to process a series of weighted inputs and generate an output in the hidden layer, which lies between the input and output layers. The values of the hidden layer nodes are computed by summing the input node values, each multiplied by its respective weight [131]. A feed-forward back propagation algorithm with Lavenberg-Marquardt function is utilized during network training since it is a comparatively quicker method. The process parameters are T_{on}, T_{off}, I_P and electrode material. The network's input layer consists of four neurons, each representing one of the EDM parameters. The output layer corresponds to the responses observed in the EDM process. To determine the optimal number of neurons in the hidden layer, a trial and error method is employed. The best architecture with a minimum MSE value is obtained with two hidden layers having five neurons for predicting the responses of the considered EDM process. Figure 5.2 shows the architecture of the FFNN that offers the maximum predictive accuracy.

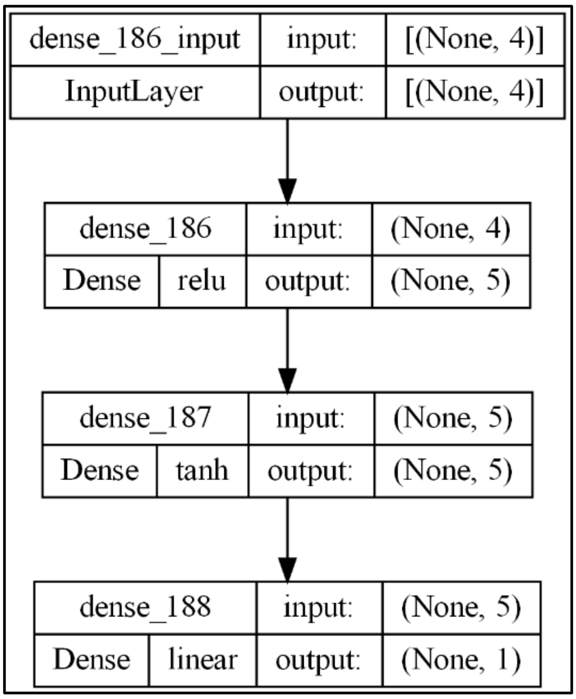


Figure 5.2 Developed FFNN architecture

A sequence of one input layer, two dense (hidden) layers and one output layer is thereby considered for developing the predictive FFNN model. This network is referred to as a 4-5-5-1 network because there are four input neurons, five hidden neurons and one output neuron. In this study, the corresponding learning parameters for the FFNN structure are kept as follows: log sigmoid activation function, a learning rate of 0.05 and a momentum constant of 0.96. The activation function transforms the output from the neurons before passing it to the next layer. Here, ReLU activation function is chosen for its widespread use in predicting continuous variables, as in regression. During training of the NN, MSE is employed as the loss function, which is the most commonly adopted error metric in regression problems. The Adam optimizer algorithm is utilized in conjunction with the loss function.

A single input layer, three hidden layers and one output layer are chosen as the architecture of the CNN. Among the hidden layers, one layer represents the convolution one-dimensional (Conv 1D) layer with 32 nodes. Another hidden layer is a flattened layer emerging from the nodes of Conv 1D layer. Each layer in the network is activated using ReLU function. The model is trained for 20 epochs with batch size ten, and further compiled with MSE loss function and Adam optimizer. The last layer is the output layer [132]. Figure 5.3 depicts the developed architecture for the CNN.

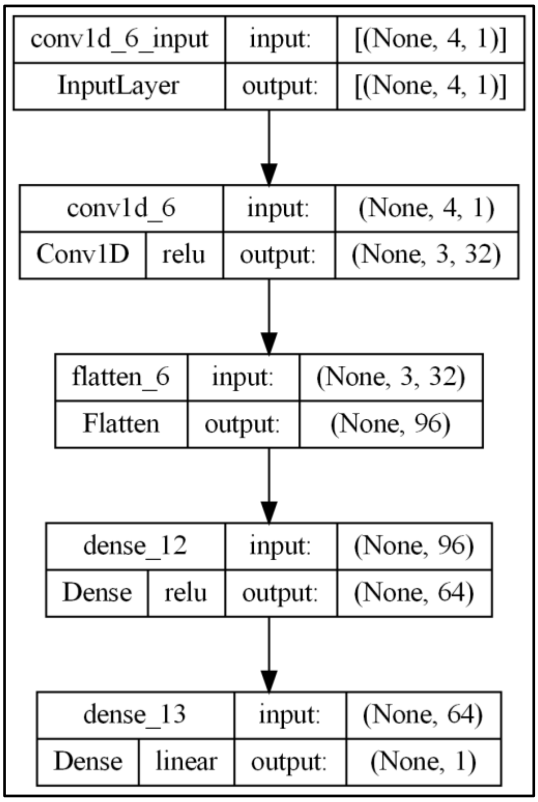


Figure 5.3 Developed CNN architecture

To develop the corresponding LSTM model, a sequence consisting of one input layer, one hidden layer and one output layer is considered. The hidden layer includes LSTM layer with 64 nodes. Neurons in the hidden layer receive signals from the input layer and transmit them to the output layer. During compilation, Adam optimizer is employed with MSE as the loss function. Figure 5.4 represents the developed LSTM architecture.

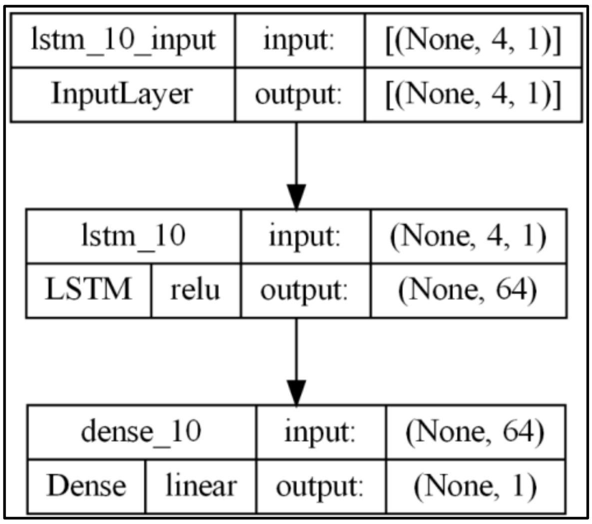


Figure 5.4 Developed LSTM architecture

RNN is developed with one input layer, one hidden layer and one output layer which follows a sequential approach by collecting series of information as inputs and processing the output as a sequence of data. The RNN layer performs the same task for every layer in a sequence, with output being dependent on the previous calculations, hence, called as the ‘recurrent’. The hidden layer in the network is represented by a simple RNN layer. Here, ReLU is utilized as the activation function and Adam-based optimization is carried out during the compilation process. The architecture derived for RNN is portrayed in Figure 5.5.

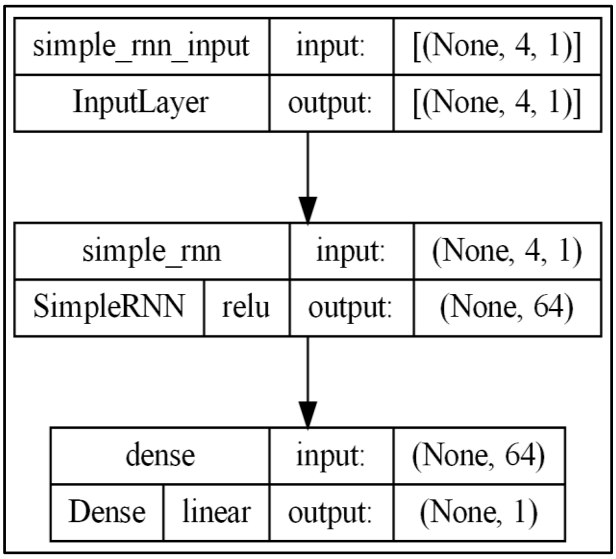


Figure 5.5 Developed RNN architecture

Four important NN-based models are developed employing specific architectures and parameters to study their prediction performance for an EDM process, considering T_{on}, T_{off}, I_P and electrode material as an input variables, and SR, MRR, TWR, OC, TC, CIR and CYL as output variables. After training, the corresponding predicted values for SR, MRR, TWR, OC, TC, CIR and CYL are obtained, as shown in Table 5.3. For more clear visualization, the scatter plots between the predicted values and actual values for all the seven responses using FFNN, CNN, LSTM and RNN are displayed in Figures 5.6-5.12, respectively. The developed models are further validated while developing those plots to demonstrate the relationship between the predicted and actual response values for the said EDM operation. In other words, the developed scatter plots display the degree of agreement between the response values predicted by the four NN models and actual experimental observations. The deviations of data points from the diagonal line are found to be minor and within the acceptable range of ±15%, relative to the average diagonal line.

Table 5.3 Actual and predicted response values for EDM responses

Exp. No.	SR					MRR				
	Actual	FFNN	CNN	LSTM	RNN	Actual	FFNN	CNN	LSTM	RNN
1	3.2	3.2708	3.8993	3.4905	3.1972	7.5512	3.1518	3.6546	12.3407	17.7488
2	6.68	6.3823	6.7194	6.7219	6.3881	61.181	56.007	38.9813	56.2995	42.4702
3	9.89	9.8635	9.7364	9.8976	9.5087	89.4419	86.316	81.9633	91.8378	60.8433
4	11.31	10.8062	10.744	10.4144	10.5894	47.7563	59.246	74.6001	48.3544	72.1793
5	5.00	5.1337	5.5104	5.4146	5.4863	34.8201	36.0123	39.0275	33.2956	14.4063
6	4.96	5.3501	6.5757	5.8945	5.9999	30.4132	36.9683	52.9695	31.8134	72.2867
7	10.02	9.7755	9.1761	9.7049	9.2349	81.8268	65.6602	54.5510	81.9531	80.3055
8	9.79	12.206	11.593	10.8914	10.282	129.2052	130.8505	126.1245	128.2356	122.224
9	6.55	5.5892	6.1355	5.3059	5.7067	14.94	13.3073	15.7569	16.2336	20.6036
10	11.27	10.455	10.4935	10.2781	10.1839	73.1567	76.5602	69.2917	74.3867	62.0047
11	11.38	11.074	10.5760	9.7121	9.3268	131.6526	140.4928	133.3954	131.9687	134.6833
12	12.76	12.137	11.8266	11.3274	11.2053	156.8343	156.8644	143.3007	156.3565	157.9921
13	4.80	5.8425	5.2981	5.1874	4.9762	46.3972	36.91805	37.6429	46.23534	34.1552
14	8.04	6.9744	8.3969	7.6826	7.8363	33.2228	52.03381	65.0566	33.23595	55.0246
15	9.00	9.6842	10.0370	9.1875	9.0996	167.2435	148.9947	121.5079	167.1308	152.9425
16	14.28	13.6783	14.4997	14.0537	13.4874	202.698	202.6873	210.7086	201.9911	197.0799
Exp. No.	TWR					OC				
	Actual	FFNN	CNN	LSTM	RNN	Actual	FFNN	CNN	LSTM	RNN
1	0.0042	0.0088	0.0018	0.0035	0.0043	0.014	0.0150	0.0171	0.0023	0.0225
2	0.0206	0.0281	0.0195	0.0239	0.0210	0.0865	0.0935	0.0466	0.0758	0.05694
3	0.0184	0.0184	0.0188	0.0207	0.0190	0.1075	0.1070	0.1145	0.1111	0.1114
4	0.0014	0.0092	-0.0078	-0.00014	0.0025	0.0835	0.0943	0.0785	0.0751	0.0875
5	0.0062	0.0133	0.0064	0.0072	0.0064	0.0715	0.0691	0.0719	0.0659	0.0729
6	0.0015	0.0054	0.0013	0.0057	0.0018	0.033	0.0305	0.0599	0.0527	0.0558
7	0.0023	-1.21E-05	-0.0042	-0.0027	0.0028	0.0625	0.05735	0.0772	0.0828	0.0756
8	0.0369	0.0322	0.0418	0.0320	0.0274	0.1	0.131021	0.1133	0.1066	0.1209
9	0.001	-0.000	-0.0004	0.0007	0.0012	0.0315	0.02112	0.0317	0.0175	0.0318
10	0.0063	0.0133	0.0083	0.0106	0.0066	0.1603	0.131639	0.1576	0.1456	0.1627

Table 5.3 Contd.

11	0.0429	0.0358	0.0353	0.0386	0.0333	0.155	0.159453	0.1767	0.1614	0.1638
12	0.0084	0.0103	0.0167	0.0105	0.0096	0.214	0.199154	0.1812	0.2160	0.1975
13	0.0255	0.0133	0.0250	0.0171	0.0258	0.01	0.053751	0.0108	0.0487	0.0120
14	0.001	0.0035	0.002	-0.0027	0.0010	0.1375	0.100865	0.1392	0.1177	0.1362
15	0.0091	0.0093	0.0090	0.0131	0.0094	0.169	0.192438	0.1538	0.1683	0.1701
16	0.0116	0.0134	0.0045	0.0117	0.0123	0.2855	0.282002	0.2890	0.2913	0.2993
Exp. No.	TC					CIR				
	Actual	FFNN	CNN	LSTM	RNN	Actual	FFNN	CNN	LSTM	RNN
1	0.061	0.0654	0.0558	0.0559	0.0599	0.0184	0.0123	0.0123	0.0184	0.0183
2	0.045	0.04203	0.0475	0.0456	0.0417	0.0115	0.0114	0.0103	0.0106	0.0095
3	0.037	0.0368	0.0296	0.0383	0.0346	0.0081	0.0105	0.0201	0.0105	0.0081
4	0.037	0.0428	0.0214	0.0410	0.0334	0.0068	0.0085	0.00016	0.0058	0.0044
5	0.01	0.0124	0.0059	0.01706	0.0079	0.0412	0.0375	0.0336	0.0416	0.0415
6	0.059	0.0576	0.0413	0.0537	0.0529	0.0074	0.0110	0.0187	0.0069	0.0222
7	0.033	0.0310	0.0283	0.0278	0.0323	0.0735	0.0731	0.0654	0.0702	0.0661
8	0.039	0.0315	0.0138	0.0364	0.0342	0.0097	0.0053	0.0032	0.0095	0.0160
9	0.002	0.0110	-0.0011	0.0102	0.00060	0.0353	0.0356	0.0167	0.0333	0.0361
10	0.023	0.0316	0.0126	0.0200	0.0193	0.0062	0.0086	0.0030	0.0050	0.0075
11	0.028	0.0271	0.0163	0.0324	0.0265	0.0456	0.0423	0.0406	0.0438	0.0397
12	0.015	0.0157	0.0011	0.01157	0.0115	0.0023	0.0015	0.0174	0.0023	0.0018
13	0.039	0.0228	0.0227	0.0287	0.0373	0.0043	0.0081	-0.0140	0.0044	0.0050
14	0.006	0.0110	-0.0041	0.0050	0.0034	0.0269	0.0277	0.0203	0.0277	0.0273
15	0.004	0.0091	-0.0066	0.0074	0.0012	0.0311	0.0285	0.0249	0.0314	0.0277
16	0.004	0.0092	-0.0112	0.0017	-0.00018	0.0645	0.0679	0.0599	0.0632	0.0629

Table 5.3 Contd.

Exp. No.	CYL				
	Actual	FFNN	CNN	LSTM	RNN
1	0.0421	0.0499	0.0396	0.0512	0.0402
2	0.0677	0.0689	0.0664	0.0551	0.0715
3	0.0905	0.0898	0.0905	0.0944	0.0874
4	0.0746	0.0693	0.0815	0.0794	0.0742
5	0.1209	0.1119	0.1171	0.1154	0.1217
6	0.0671	0.0637	0.0581	0.0638	0.0557
7	0.0749	0.0674	0.0827	0.0739	0.0782
8	0.0763	0.0761	0.0753	0.0739	0.0786
9	0.064	0.0467	0.0587	0.0619	0.0654
10	0.0814	0.0903	0.0773	0.0888	0.0825
11	0.0454	0.0774	0.0606	0.0527	0.0593
12	0.0921	0.0783	0.0886	0.0988	0.0836
13	0.0452	0.0541	0.03558	0.0502	0.0445
14	0.0628	0.0602	0.0559	0.0632	0.0666
15	0.078	0.0755	0.0786	0.0777	0.0768
16	0.1275	0.1153	0.1301	0.1233	0.1213

These near-linear patterns of the graphs indicate normally distributed datasets, with subsequent development of accurate models suitable for predictive purposes. Figures 5.6-5.12 demonstrate that all the developed NN models demonstrate a strong correlation between the actual and predicted response values.

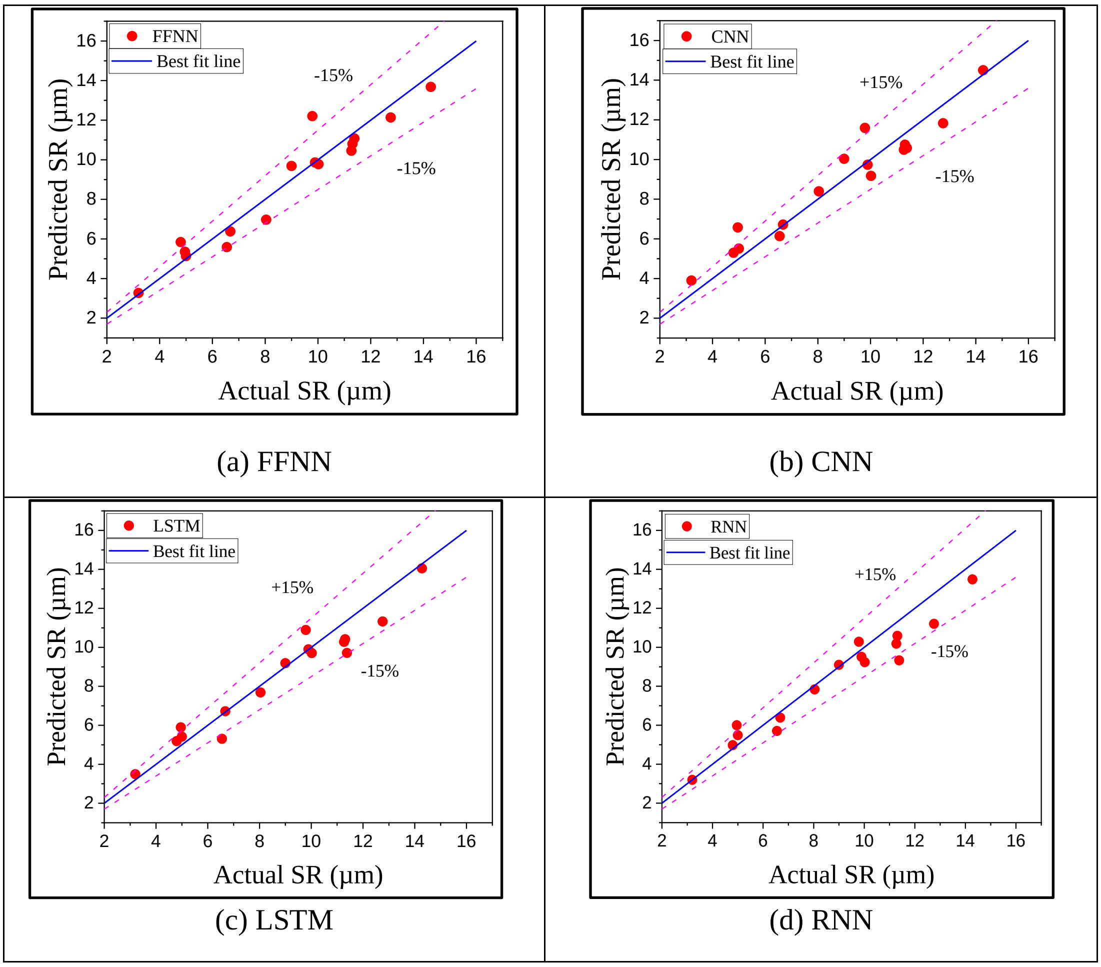


Figure 5.6 Comparison between actual and predicted values for SR by FFNN, CNN, LSTM, and RNN

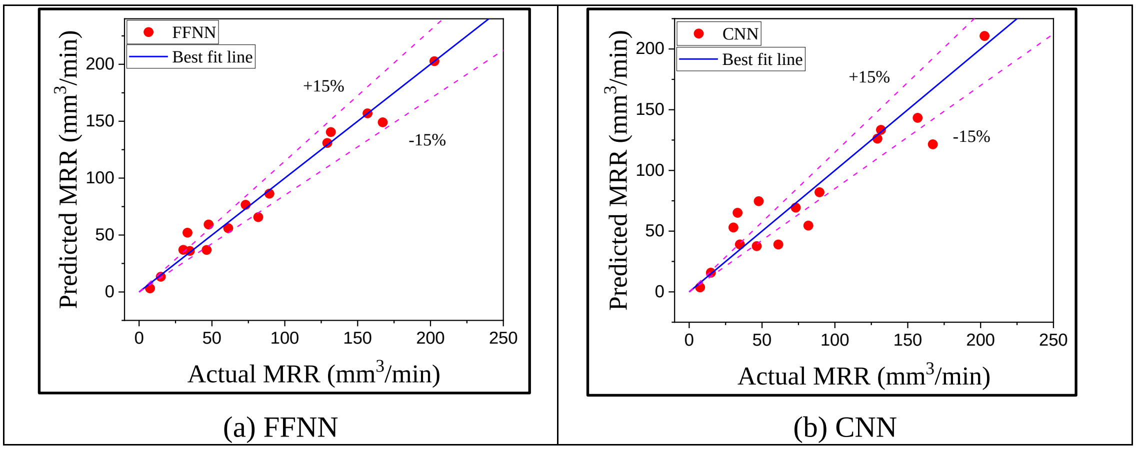


Figure 5.7 Comparison between actual and predicted values for MRR by FFNN, CNN, LSTM and RNN

But, it is important to mention that while analyzing these graphs, it can be noticed that SR and MRR have slightly more variations in their predicted results. The developed NN

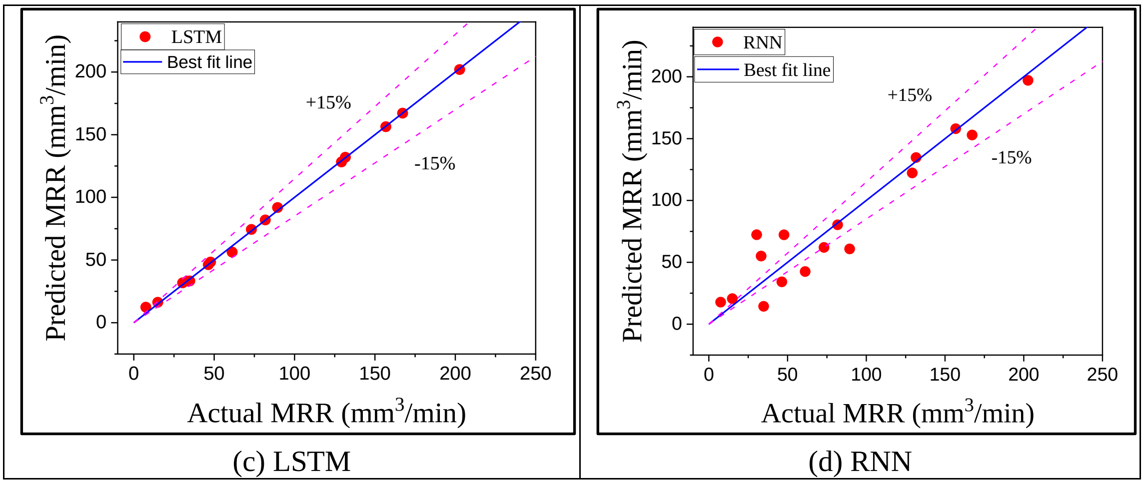


Figure 5.7 Contd.

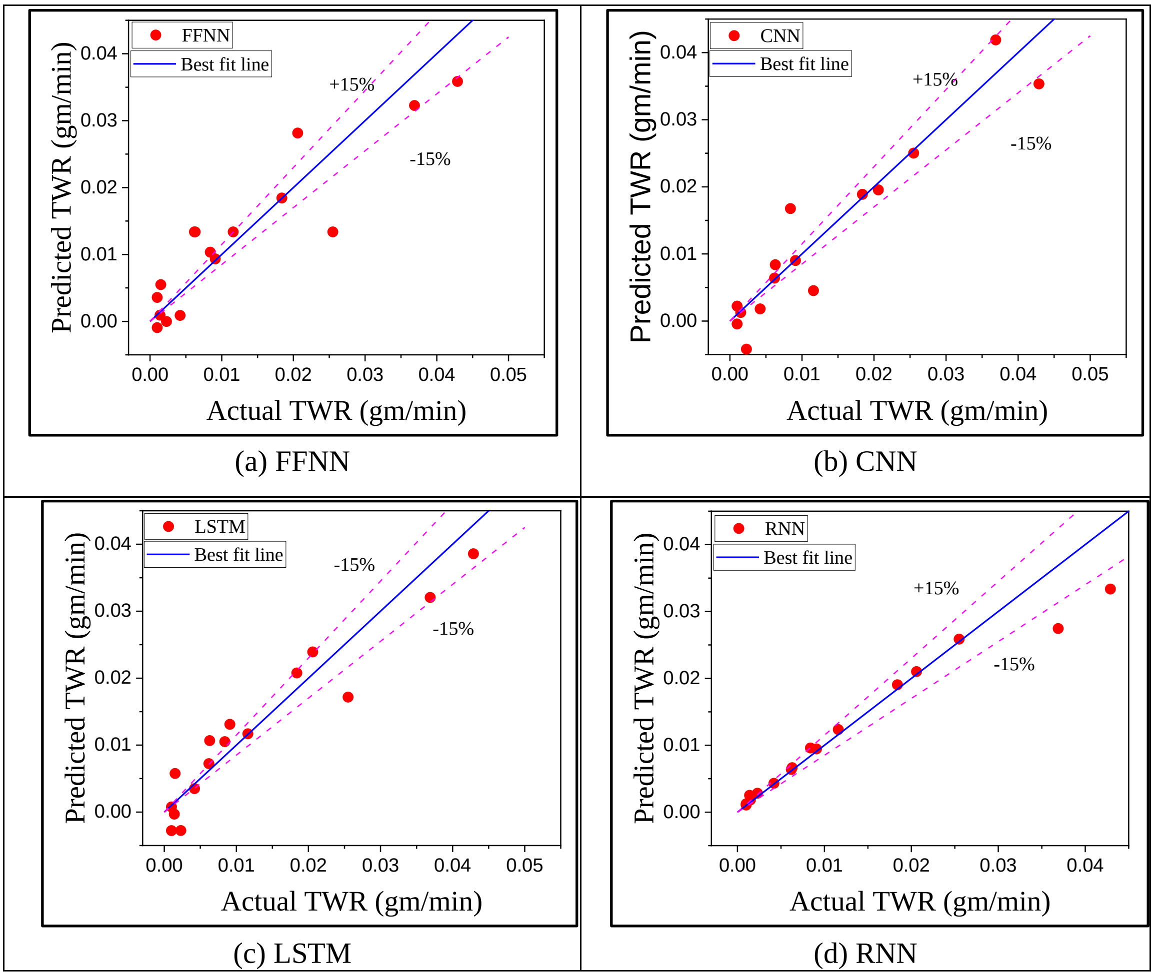


Figure 5.8 Comparison between actual and predicted values for TWR by FFNN, CNN, LSTM and RNN

models are subsequently tested to analyze and compare their prediction performance. The comparison of the experimental results and the predicted values for the corresponding NN models are shown in Figure 5.13. The proximity of data points to the diagonal line suggests that the residuals are nearly linear (normally distributed), despite some inherent randomness in the error component. This figure also shows that all the developed NN models are quite capable of predicting the considered responses within ±15% error ranges for the said EDM process.

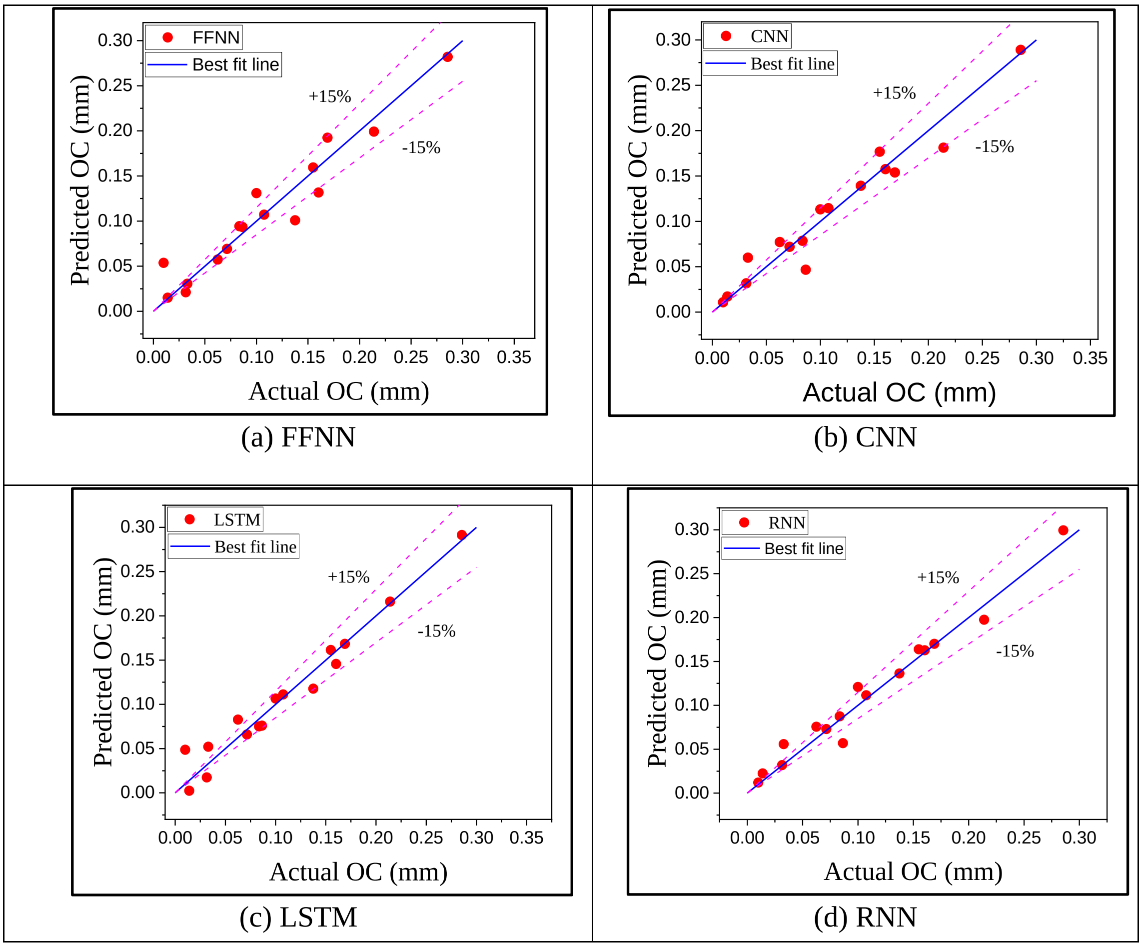


Figure 5.9 Comparison between actual and predicted values for OC by FFNN, CNN, LSTM and RNN

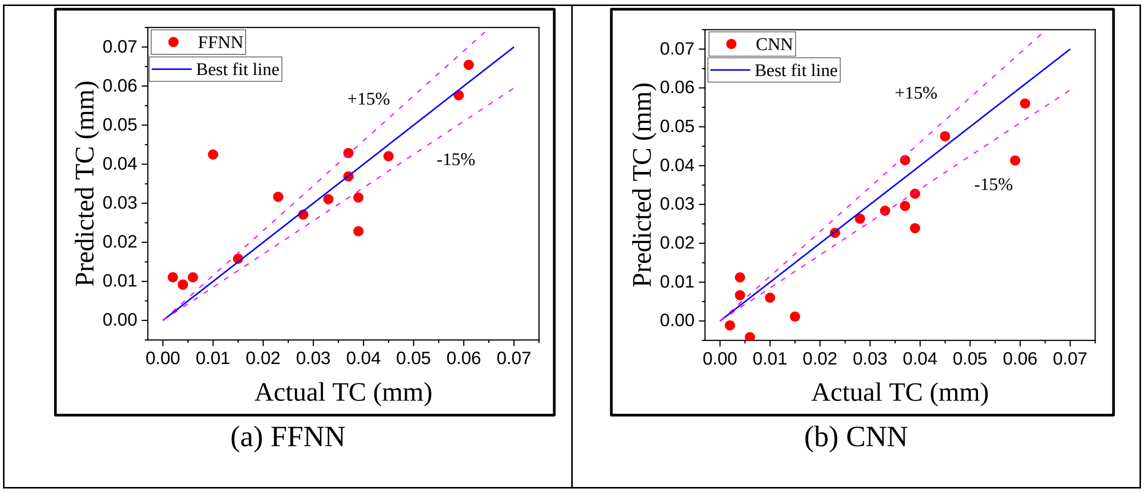


Figure 5.10 Comparison between actual and predicted values for TC by FFNN, CNN, LSTM and RNN

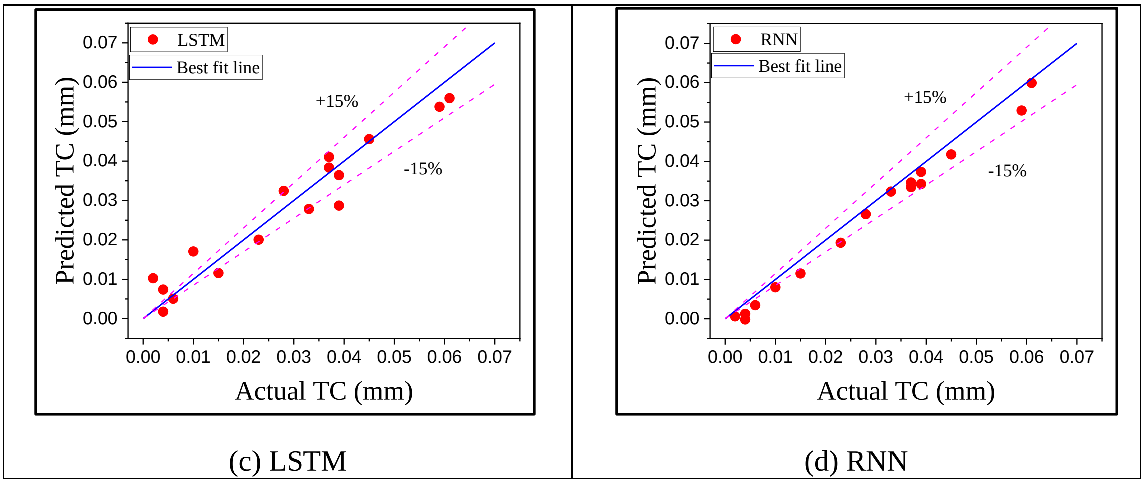


Figure 5.10 Contd.

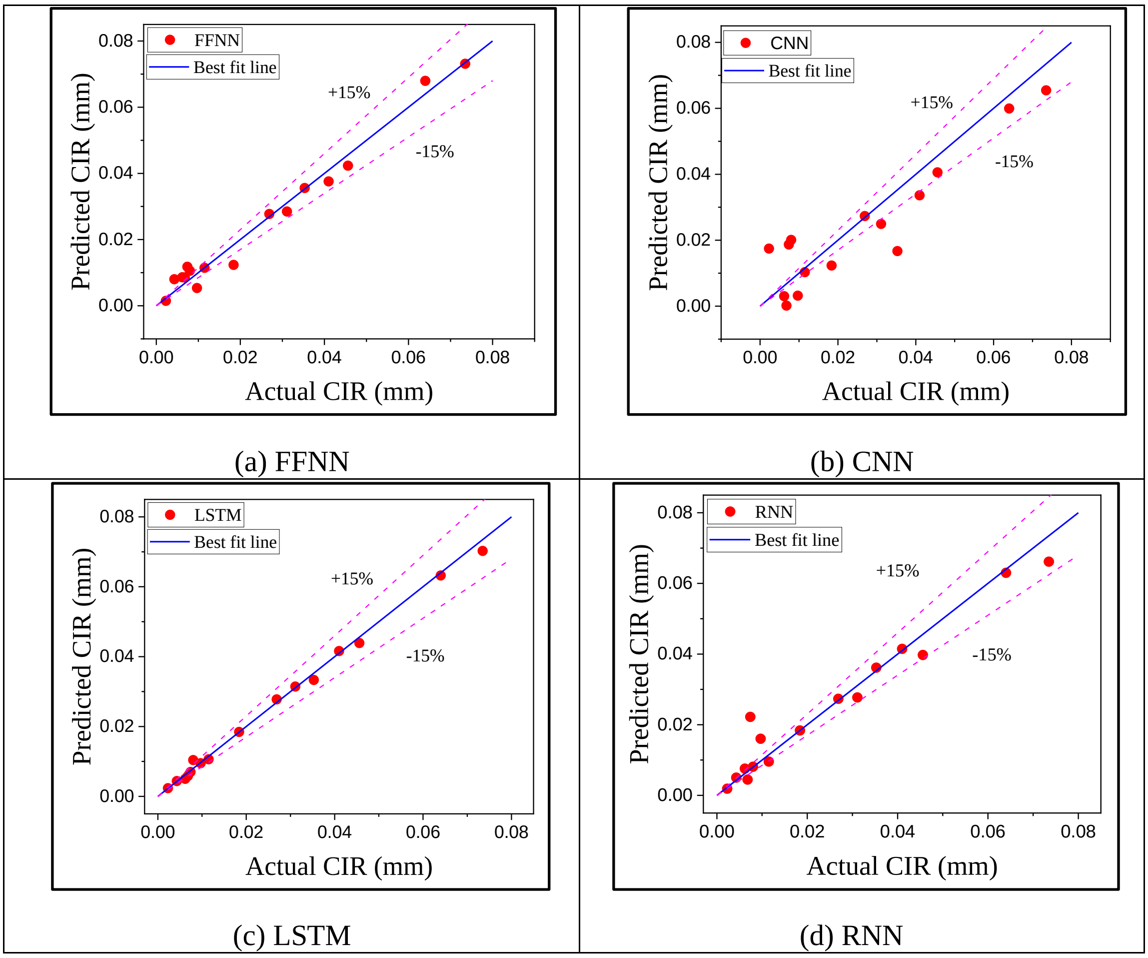


Figure 5.11 Comparison between actual and predicted values for CIR by FFNN, CNN, LSTM and RNN

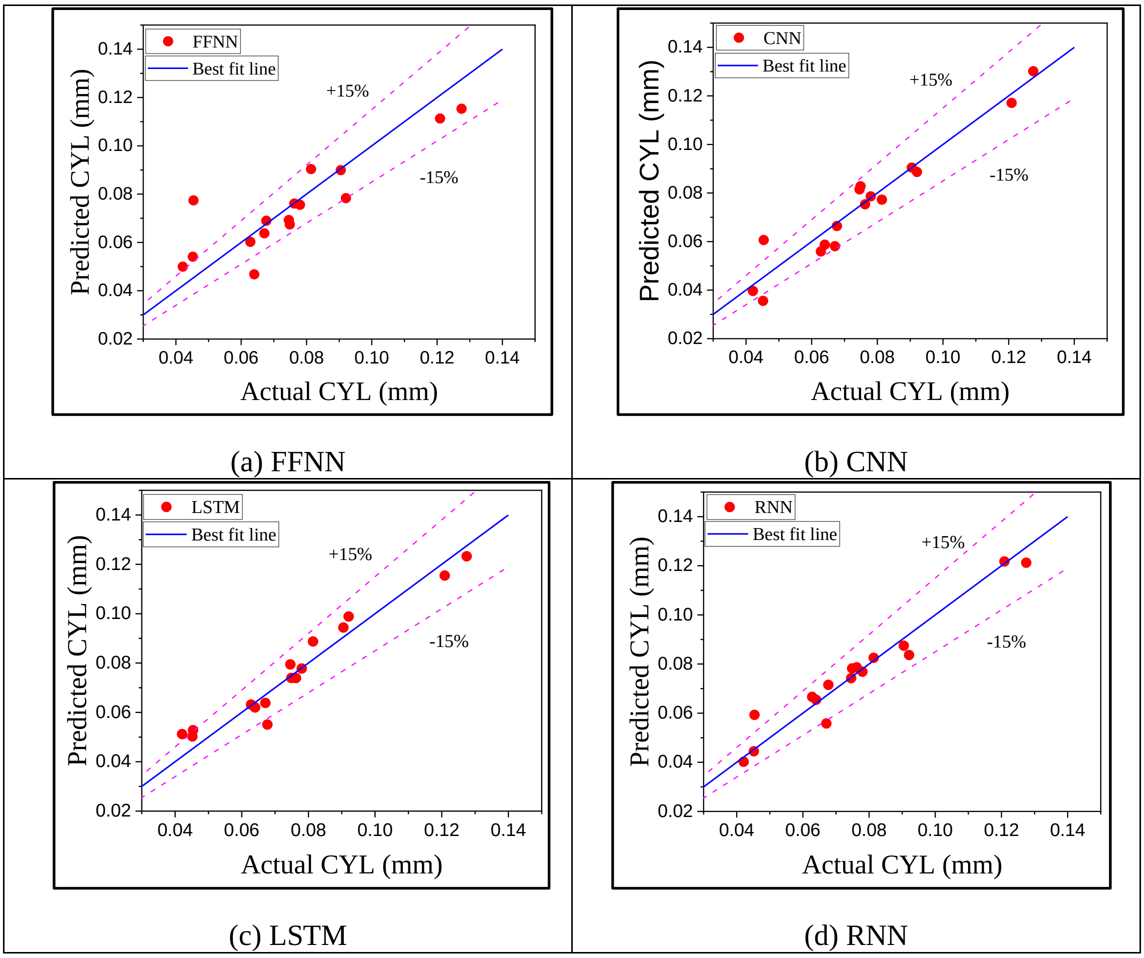


Figure 5.12 Comparison between actual and predicted values for CYL by FFNN, CNN, LSTM and RNN

In case of SR, there are good agreements between the actual and predicted values. Nearly, all the SR values predicted by the various NN models fall within the ± 15% error range of the actual values. For CYL, a similar trend can also be observed. For MRR prediction, it can be revealed that FFNN and LSTM are nearly ideal as most of the data points align closely with the diagonal line. However, in case of CNN and RNN models, few data points are noticed to be beyond the ±15% range.

In case of TC and CIR, closer correlation between the predicted and actual values can be observed for all the NN models, except CNN. On the other hand, for TWR, best predictions are shown by both RNN and LSTM as compared to FFNN and CNN, which exhibit relatively poor prediction, particularly at lower TWR values. The similarity between the datasets and pattern of the plots indicates that the predicted values are closer to their actual values. This, in turn, further decides the accuracy and effectiveness of the NN models for correlative and predictive purposes.

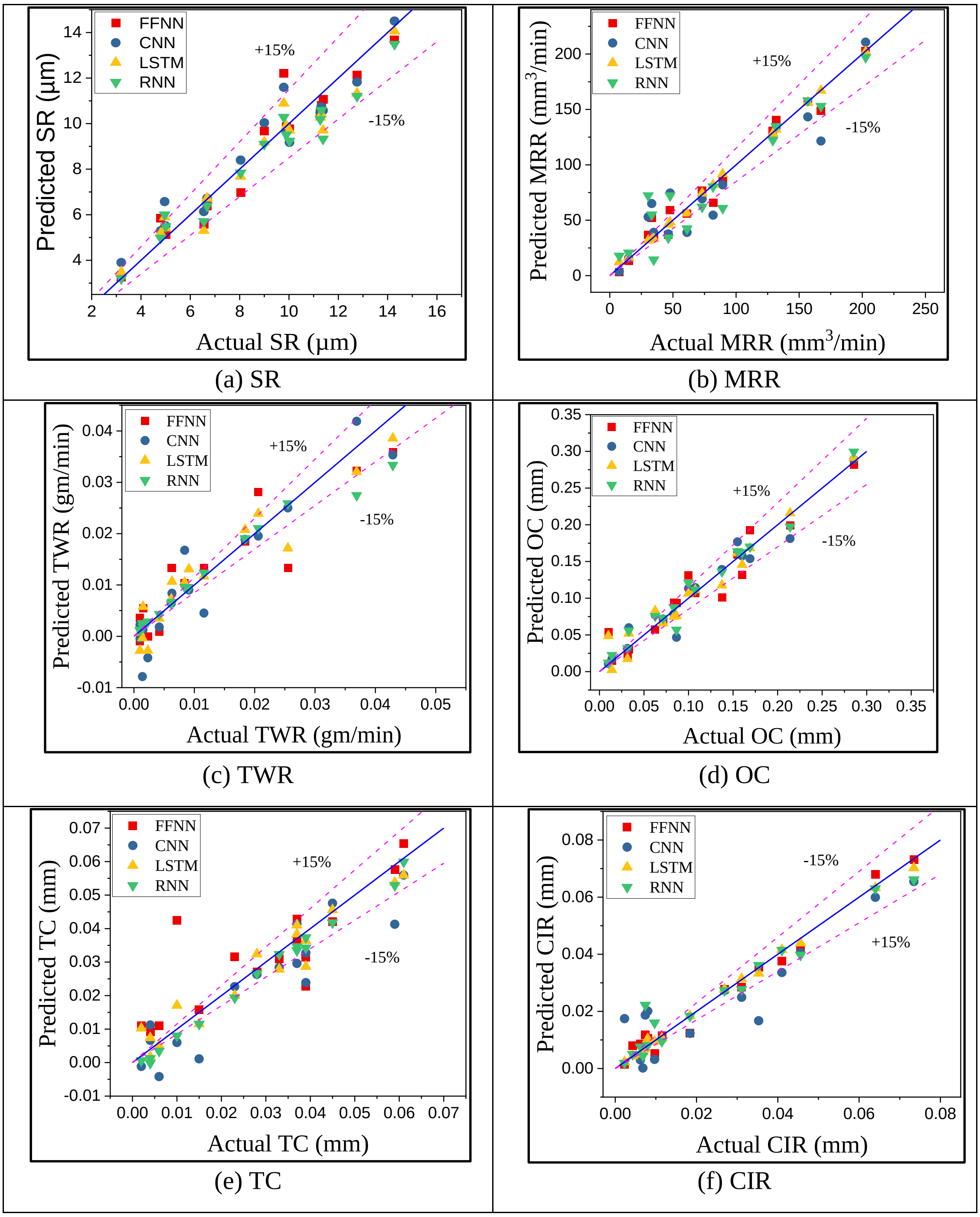
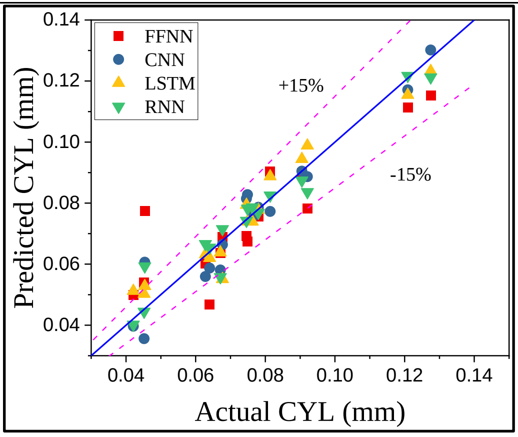


Figure 5.13 Comparison between actual and predicted values for FFNN, CNN, LSTM and RNN



(g)

Figure 5.13 Contd.

To check and further analyze predictive performance of the adopted NN models, their residual plots are depicted in Figure 5.14. In these plots, the horizontal line represents zero error in the prediction. Points lying above and below it indicate underprediction (target values greater than the predicted values) and overprediction (target values lower than the predicted values). The proximity of the data to the horizontal line suggests that the residuals are approximately linear and normally distributed, though there is still some random variability. The results also portray that all the four NN models can generalize well between the input and output variables, providing reasonably accurate predictions. In general, performance of all the NN models shows satisfactory prediction indicating lack of biasness, but it is important to note that few data points in case of SR and MRR show some scatter of the residuals. The CNN is found to have extreme variation from the normality assumption.

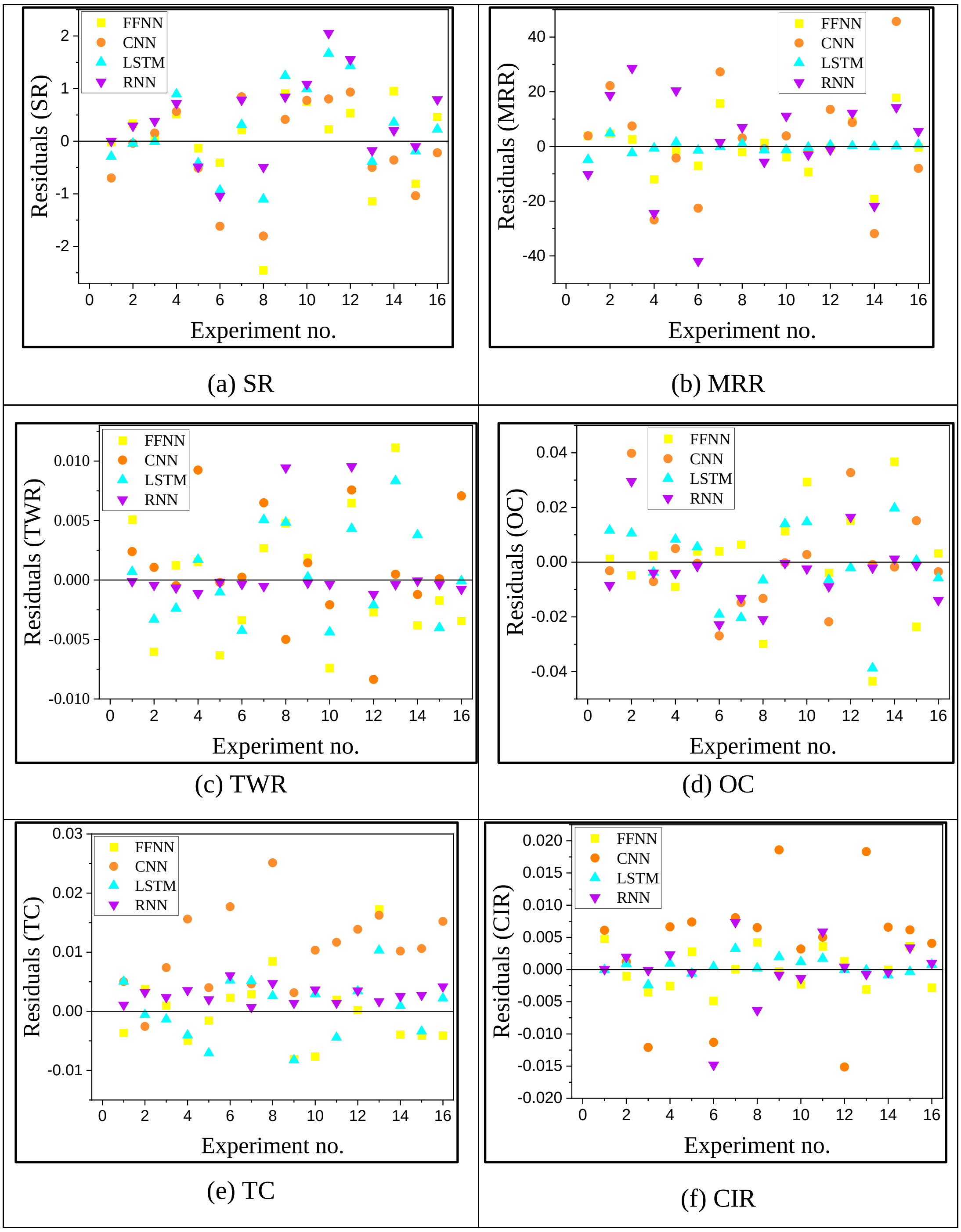
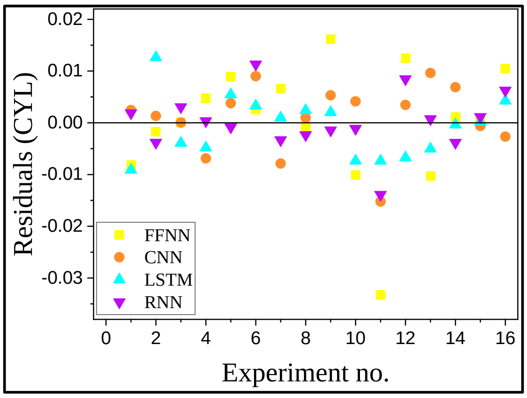


Figure 5.14 Residual plots of predicted responses by various NN models for SR, MRR, TWR, OC, TC, CIR and CYL



(g) CYL

Figure 5.14 Contd.

In order to further validate prediction performance of the four developed NN models, values of R^2, R^2_{adj}, MSE and RMSE for testing, training and overall datasets are computed, as depicted in Table 5.4. The MSE measures the average squared difference between the predicted outputs and actual targets. Lower MSE values indicate smaller, less significant errors, suggesting better performance of the developed models. RMSE is a metric used to measure accuracy of a model's prediction. RMSE provides a way to quantify the magnitude of error in a model, with lower values indicating better performance. It is important to mention here that for effective prediction by NN models, maximum values of R^2 and R^2_{adj}, and minimum values of MSE and RMSE are always desirable. As observed from Table 5.4 and Figure 5.13, SR values obtained during testing of the NN models are closer to the experimental results. Based on the overall database, the R^2 values are noticed to be 0.9343, 0.9111, 0.9294 and 0.9215 for FFNN, CNN, LSTM and RNN, respectively. The R^2_{adj} values are also more than 87% for all the four NN models. The derived results for SR reveal that FFNN, LSTM and RNN show good prediction performance with higher R^2 and R^2_{adj} values, and correspondingly lower values of MSE and RMSE based on the overall dataset as compared to CNN. Even though CNN provides satisfactorily higher values for R^2 and R^2_{adj} (0.9111 and 0.8785) for SR, but the MSE and RMSE show higher values (0.8631 and 0.9392) as compared to other models. Thus, LSTM and RNN prove to be the best models, followed by FFNN in predicting SR values during the considered EDM operation.

While predicting MRR with the training dataset, LSTM demonstrates the best performance, achieving the highest R^2 and R^2_{adj} values of 0.9903 and 0.9982, respectively. It is further supported by its lowest MSE and RMSE values of 4.0300 and 2.0063, respectively. LSTM is also emerged to be the best in predicting MRR values using the testing data with R^2, R^2_{adj}, MSE and RMSE values as 0.9908, 0.9985, 3.0908 and 1.7573, respectively. Considering both the training and testing datasets together, LSTM again evolves out as the best NN model with corresponding R^2, R^2_{adj}, MSE and RMSE values as 0.9975, 0.9856, 3.8915 and 1.9810, respectively. Based on the obtained values of R^2, R^2_{adj}, MSE and RMSE for prediction of MRR, FFNN occupies the second best position, followed by RNN and CNN.

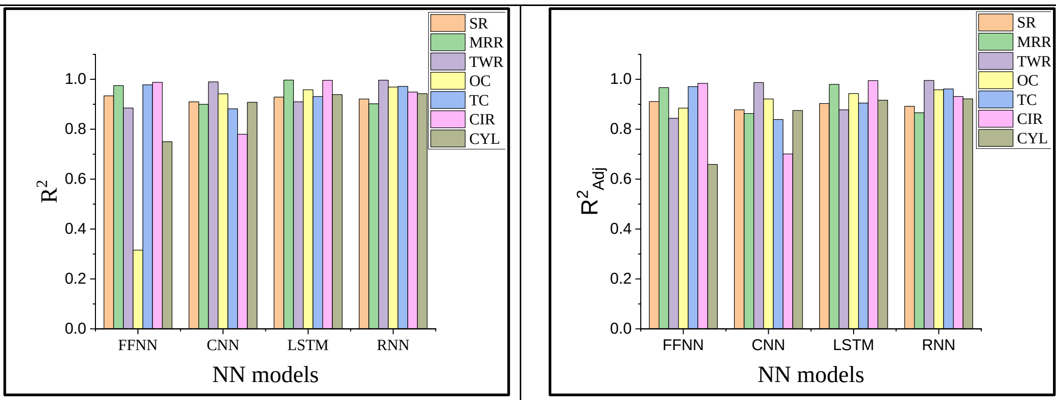
While considering TWR, the maximum values of R^2, R^2_{adj}, and minimum values of MSE and RMSE with respect to overall dataset as 0.9968 and 0.9956, and 3.26E-07 and 0.00057, respectively are obtained from the developed RNN model. The corresponding values obtained from CNN for TWR are also very close to the prediction performance values obtained for RNN.

Table 5.4 Calculated statistical metrics for the FFNN, CNN, LSTM and RNN models

Response	Model	Dataset	R^2	R^2_{adj}	MSE	RMSE
SR	FFNN	Test	0.9141	0.883	0.7261	0.8522
		Train	0.9411	0.9210	0.5912	0.7694
		Overall	0.9343	0.9110	0.6324	0.7953
	CNN	Test	0.8862	0.8580	0.9152	0.9566
		Train	0.9373	0.8921	0.8541	0.9248
		Overall	0.9111	0.8785	0.8631	0.9392
	LSTM	Test	0.9177	0.8971	0.6668	0.8158
		Train	0.9300	0.9129	0.6905	0.8304
		Overall	0.9294	0.9031	0.6869	0.8282
	RNN	Test	0.9142	0.8920	0.7476	0.8641
		Train	0.9223	0.9030	0.7634	0.8731
		Overall	0.9215	0.8920	0.7608	0.8745
MRR	FFNN	Test	0.9734	0.9632	89.7025	9.4745
		Train	0.9772	0.9684	74.4456	8.6206
		Overall	0.9751	0.9673	79.0012	8.8802
	CNN	Test	0.8989	0.8732	319.0312	17.841
		Train	0.8990	0.8741	328.2502	18.1251
		Overall	0.9001	0.8631	327.1501	18.0824
	LSTM	Test	0.9908	0.9985	3.0908	1.7573
		Train	0.9903	0.9982	4.0300	2.0063
		Overall	0.9975	0.9856	3.8915	1.9810
	RNN	Test	0.9061	0.8832	2.91E+02	17.0600
		Train	0.9057	0.8764	3.27E+02	18.0800
		Overall	0.9023	0.8663	3.20E+02	17.8841
TWR	FFNN	Test	0.9051	0.8714	1.54E-05	0.0039
		Train	0.8772	0.8322	1.97E-05	0.0443
		Overall	0.8854	0.8441	1.84E-05	0.0043
	CNN	Test	0.9956	0.9881	1.41E-06	0.0012
		Train	0.9932	0.9880	1.49E-06	0.0012
		Overall	0.9902	0.9870	1.4822	0.0012
	LSTM	Test	0.896	0.8710	1.61E-05	0.0040
		Train	0.9121	0.8920	1.41E-05	0.004
		Overall	0.9185	0.8784	1.44E-05	0.0038
	RNN	Test	0.9965	0.9957	3.44E-07	0.0006
		Train	0.9969	0.9961	3.21E-07	0.0005
		Overall	0.9968	0.9956	3.26E-07	0.0005

Table 5.4 Contd.

Response	Model	Dataset	R^2	R^2_{adj}	MSE	RMSE
OC	FFNN	Test	0.9031	0.868	4.81E-04	0.0219
		Train	0.9211	0.8912	4.40E-04	0.0211
		Overall	0.3155	0.8848	4.57E-04	0.0213
	CNN	Test	0.9371	0.9223	3.30E-04	0.0182
		Train	0.9421	0.9283	3.09E-04	0.0175
		Overall	0.9424	0.9214	3.13E-04	0.0177
	LSTM	Test	0.9561	0.9452	2.32E-04	0.0152
		Train	0.9590	0.9491	2.20E-04	0.0148
		Overall	0.9582	0.9436	2.20E-04	0.0149
	RNN	Test	0.9694	0.9622	1.61E-04	0.0127
		Train	0.9695	0.9613	1.67E-04	0.0129
		Overall	0.9698	0.9584	1.66E-04	0.0129
TC	FFNN	Test	0.9794	0.9712	7.24E-06	0.0027
		Train	0.9785	0.9700	7.44E-06	0.0027
		Overall	0.9785	0.9710	7.38E-06	0.0027
	CNN	Test	0.8655	0.8322	3.77E-05	0.0061
		Train	0.8844	0.8555	4.17E-05	0.0064
		Overall	0.882	0.8392	4.11E-05	0.0064
	LSTM	Test	0.9163	0.8963	2.33E-05	0.0048
		Train	0.9328	0.9164	2.42E-05	0.0049
		Overall	0.9312	0.9054	2.41E-05	0.0049
	RNN	Test	0.9678	0.9589	9.54E-06	0.0031
		Train	0.9720	0.9651	9.94E-06	0.0031
		Overall	0.9717	0.9615	9.86E-06	0.0031
CIR	FFNN	Test	0.9874	0.9828	5.33E-06	0.0023
		Train	0.9887	0.9841	5.38E-06	0.0023
		Overall	0.9885	0.9844	5.36E-06	0.0023
	CNN	Test	0.8077	0.7592	9.87E-05	0.0099
		Train	0.7753	0.7192	1.01E-04	0.0112
		Overall	0.7801	0.7010	1.00E-04	0.0112
	LSTM	Test	0.9964	0.9958	1.95E-06	0.0014
		Train	0.9963	0.9954	1.73E-06	0.0013
		Overall	0.9961	0.9947	1.77E-06	0.0013
	RNN	Test	0.9621	0.9532	1.88E-05	0.0043
		Train	0.9450	0.9323	2.44E-05	0.0049
		Overall	0.9495	0.9312	2.33E-05	0.0048
CYL	FFNN	Test	0.7488	0.6460	1.22E-04	0.011
		Train	0.7537	0.6630	1.40E-04	0.0118
		Overall	0.7522	0.6591	1.35E-04	0.0116
	CNN	Test	0.8923	0.8654	5.07E-05	0.0071
		Train	0.9114	0.8891	4.92E-05	0.0070
		Overall	0.9082	0.8753	4.94E-05	0.0070
	LSTM	Test	0.9334	0.9167	3.13E-05	0.0055
		Train	0.9396	0.9245	3.34E-05	0.0058
		Overall	0.9388	0.9166	3.31E-05	0.0057
	RNN	Test	0.9464	0.9324	2.80E-05	0.0053
		Train	0.9415	0.9275	3.18E-05	0.0056
		Overall	0.9427	0.9219	3.10E-05	0.0055



(a) R^2 (b) R^2_{adj}

Figure 5.15 Prediction performances of various NN models with respect to R^2 and R^2_{adj}

For LSTM, the R^2, R^2_{adj}, MSE and RMSE values are calculated as 0.9185, 0.8784, 1.44E-05 and 0.0038, respectively, whereas, FFNN shows their corresponding values as 0.8854, 0.8441, 1.84E-05 and 0.0043, respectively. Thus, it can be observed that for TWR, RNN shows the best prediction results, followed by CNN, LSTM and FFNN. Here, it is important to highlight that CNN provides satisfactory values for R^2 and R^2_{adj} as 0.9902 and 0.9870, respectively, but accompanying with higher MSE (1.4822). However, it is important to note that the sequence of the ranking order of the developed NN models is slightly different from that obtained for MRR.

While predicting OC values, it can be noticed that RNN provides higher R^2 and R^2_{adj} values as 0.9698 and 0.9584, and lower MSE and RMSE values as 1.66E-04 and 0.0129, respectively. For LSTM, the R^2, R^2_{adj}, MSE and RMSE values are calculated as 0.9582, 0.9436, 2.20E-04 and 0.0149, respectively. Thus, it can be noticed from Table 5.4 that RNN shows the best performance with respect to all the model accuracy metrics, followed by LSTM, CNN and FFNN with marginal differences. But it should be mentioned that FFNN provides poor prediction for OC since it calculates comparatively lower value for R^2 as 0.3155.

Taking into account the test data, in case of TC, the best prediction performance results are obtained from FFNN model. This model calculates the highest R^2 and R^2_{adj} values as 0.9794 and 0.9712, respectively, while the lowest MSE and RMSE values are observed as 7.24E-06 and 0.0027 respectively. The next best model possessing superior prediction accuracy after FFNN is RNN since it shows the corresponding values with marginal differences, i.e. R^2 and R^2_{adj} values as 0.9678 and 0.9589, while MSE and RMSE as 9.54E-06 and 0.0031, respectively. Similar outputs can also be noticed while considering the overall dataset. It is important to note that in case of TC, all the NN models provide good prediction results. However, while deciding the ranking order of their performance, it is noticed that

FFNN performs best, while RNN has marginally better accuracy, followed by LSTM and CNN.

It can also be revealed from Table 5.4 and Figure 5.15 that in case of CIR, on the basis of R^2 and R^2_{adj} values, for the testing data, LSTM occupies the best position with R^2 value of 0.9964 and R^2_{adj} value of 0.9958. The values of MSE and RMSE also validate the same results. The similar result can also be confirmed by considering the overall dataset. After LSTM, FFNN shows satisfactory prediction for CIR, with values of R^2 and R^2_{adj} as 0.9885 and 0.9844, and minimum MSE and RMSE values as 5.36E-06 and 0.00231, respectively. On the other hand, R^2, R^2_{adj}, MSE and RMSE values for RNN are calculated as 0.9495, 0.931, 2.33E-05 and 0.0048, respectively. It is observed that CNN exhibits comparatively poor prediction with R^2, R^2_{adj}, MSE and RMSE as 0.7801, 0.7010, 1.00E-04 and 0.0112, respectively. Thus, it is obvious that for predicting CIR values, LSTM supersedes providing highest accuracy, followed by FFNN, RNN and CNN. It is interesting to note that the same sequence of ranking is previously achieved while comparing their prediction performance for MRR. Finally, while considering CYL, the derived values of R^2, R^2_{adj}, MSE and RMSE as 0.9427, 0.9219, 3.10E-05 and 0.0055, respectively reveal that RNN shows the maximum prediction accuracy as compared to FFNN, CNN and LSTM for the overall dataset. LSTM also provides quite satisfactory results with R^2 and R^2_{adj} as 0.9388 and 0.9166, and the corresponding values of MSE and RMSE values as 3.31E-05 and 0.0055, respectively. The results obtained from Table 5.4, and Figures 5.16-5.17 show that RNN proves to be the best prediction model, followed by LSTM, CNN and FFNN. The obtained values of R^2 (0.7522), R^2_{adj} (0.6591), MSE (1.35E-04) and RMSE (0.0116) from FFNN model assure the ranking order stated above while analyzing the prediction performance for CYL.

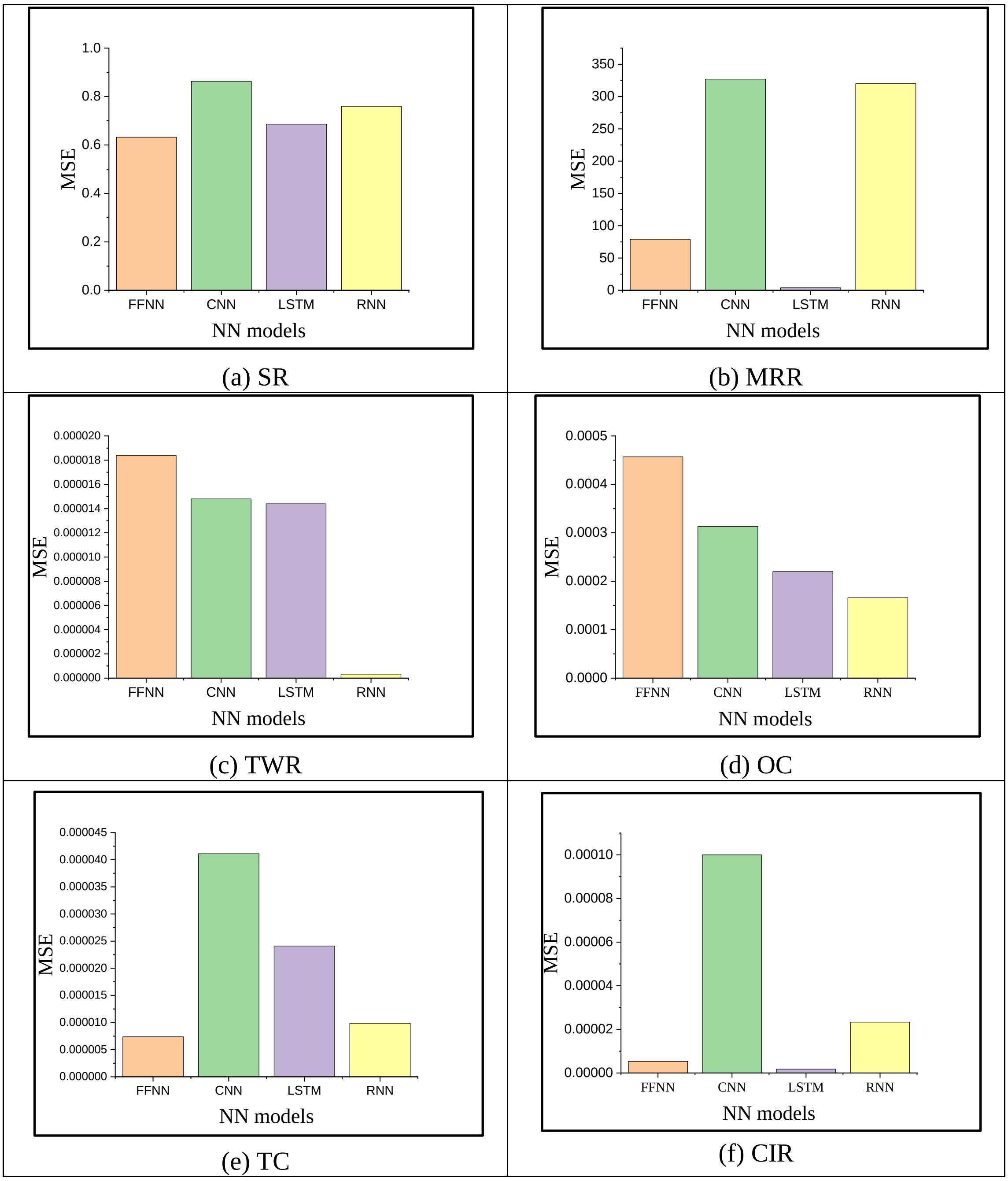
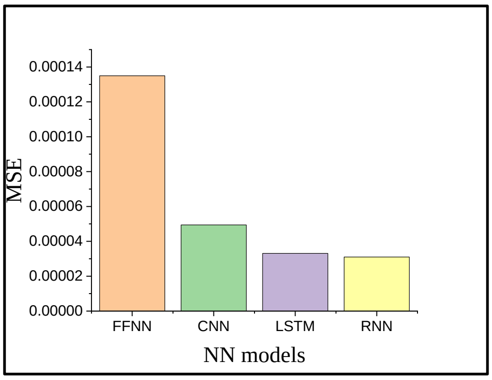
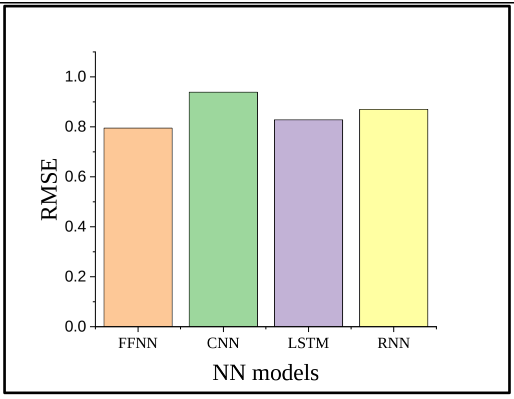


Figure 5.16 Prediction performances of the NN models for SR, MRR, TWR, OC, TC, CIR and CYL with respect to MSE

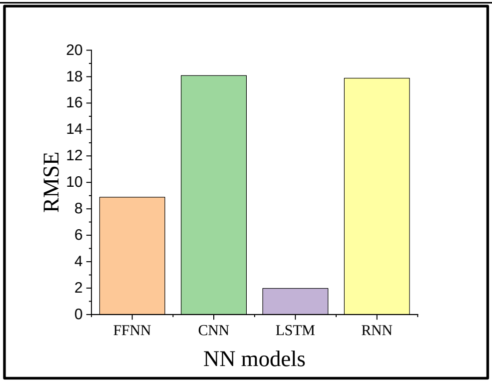


(g) CYL

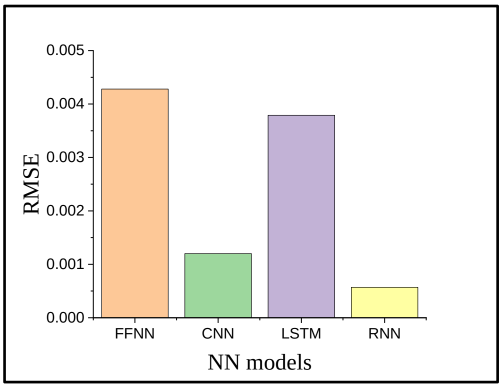
Figure 5.16 Contd.



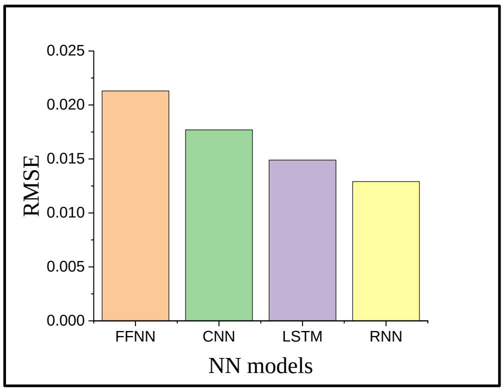
(a) SR



(b) MRR

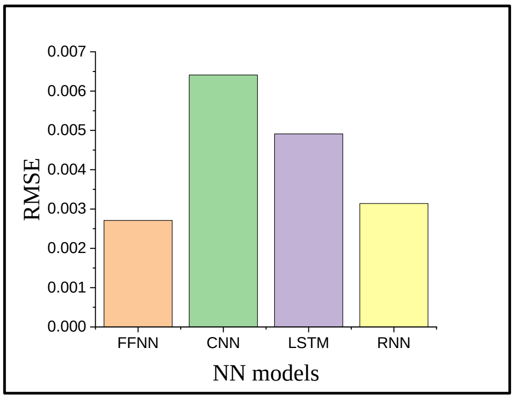


(c)TWR

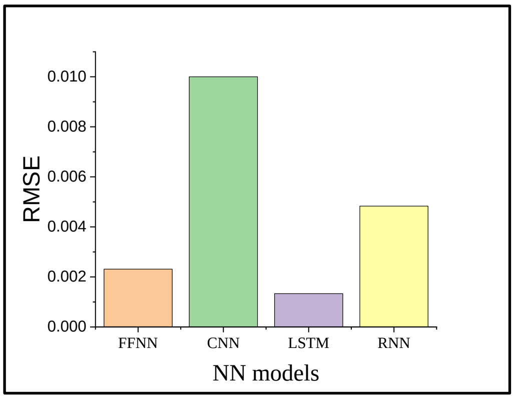


(d) OC

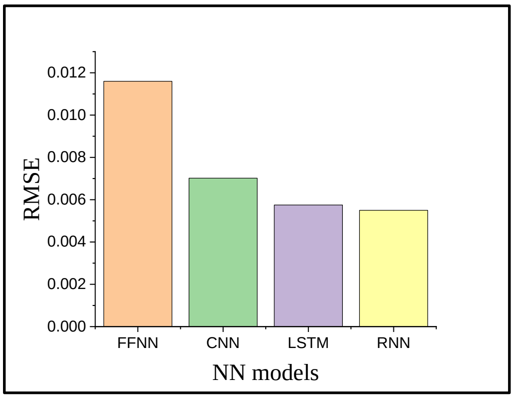
Figure 5.17 Prediction performance of the NN models for SR, MRR, TWR, OC, TC, CIR and CYL with respect to RMSE



(e) TC



(f) CIR



(g) CYL

Figure 5.17 Contd.

CONCLUSIONS

6. CONCLUSIONS

In this research work, performance of various DM tools in the form of development of DTs using CART and CHAID, NN approaches, and predictive modeling techniques, such as XGB, KRG, RBF and GEP, is analyzed and compared while solving various machining parametric analysis and optimization problems. Based on the extensive analysis of their applicability, potentiality and effectiveness, the following conclusions can be drawn:

a) In the CNC turning process, while determining the most preferable combination of various input parameters using DTs, CART algorithm supersedes CHAID algorithm with respect to higher overall classification accuracy and lower prediction risk. Although, for some of the responses, CART generates slightly more number of decision rules as compared to CHAID, it can almost perfectly predict high as well as low values of all the responses. Between these two algorithms, CART has a better capacity to predict higher values of the responses, whereas, CHAID performs better for lower response values.

b) Based on the detailed analysis of the developed decision rules, it can be observed that for achieving higher TL, cryogenic environment, and lower values of CS, NR and FR need to be set. DOC has almost no effect on TL. On the other hand, a combination of cryogenic environment, low or medium CS, and low FR and NR is responsible for lower PC.

c) The decision rules extracted from the DTs also identify that to attain lower SR of the turned components, cryogenic environment, high CS and NR, and low or medium FR, can be the recommended settings for the said CNC turning process. DOC plays an insignificant role on SR. A parametric mix of cryogenic environment, low CS and DOC, and low or medium FR and NR would provide lower CF.

d) Based on the sensitivity analysis study, the prediction performance of CART algorithm is observed to be least affected by the perturbed response values in the experimental dataset against CHAID algorithm.

e) While analyzing CART and CHAID algorithms for the WEDM process, it is observed that CART again outperforms CHAID with respect to both overall classification accuracy and prediction risk. The results also show that T_{on} is the most important parameter affecting almost all the responses, followed by T_{off} and I_P. On the other hand, WF and WT contribute least towards attainment of the target response values.

f) On the basis of the decision rules extracted from the DTs for the WEDM process, it can be derived that for achieving higher MR, high values of T_{on} and I_P, and low values of T_{off} and SV need to be maintained. On the other hand, to attain lower SR of the machined components, low or medium T_{on}, high T_{off} and SV, and low or medium I_P are recommended. WF and WT play no significant roles on SR.

g) It is observed that for the WEDM process, a mix of low or medium T_{on}, high T_{off} and SV, and small I_P is responsible for attainment of lower DD. Low T_{off}, high or medium T_{on}, high I_P and low or medium SV are accountable for lower WWR.

h) On the basis of the developed DTs and decision rules, it is observed that PR has the maximum contribution in controlling the responses during the USM process.

i) While analyzing the comparative performance of the predictive models, XGB, KRG and RBF show the best accuracy with maximum R^2 and R^2_{adj}, and minimum MSE and RMSE values. On the other hand, GEP exhibits the worst values of all the four metrics for all the considered responses. Hence, based on the statistical metrics and overall datasets, XGB, KRG and RBF outperform GEP in predicting the considered response values during the EDM process.

j) When different NN models are developed in the form of FFNN, CNN, RNN and LSTM for the EDM process, and detailed comparative analysis of their prediction performance is carried out, it is revealed that all the developed NN models demonstrate strong correlation between the experimental and predicted values of the considered responses. Thus, when the prediction performance is critically evaluated using the overall dataset, their ranking order can be derived as LSTM-RNN-FFNN-CNN with respect to the considered performance metrics as evidenced by the obtained values.

The limitations of this research work include:

a) Each algorithm possesses unique features applicable to various problems, making it impractical to determine the best DM tool for the entire machining domain.

b) As the decision rules mainly focus on classification, they often neglect predicting the interrelationships between the input parameters and responses in the form of regression. While a continuous variable is divided into intervals and turned into a classification problem, there is high possibility of loss of valuable information.

c) Large sized DTs can become opaque making it challenging in terms of visibility and clarity.

This work proposes the following future research directions:

a) As a future scope, application potentiality of other DM techniques, including RF, SVM, RST, GRNN, ELSVM etc. can be explored and their prediction performance can be compared for accurately predicting the corresponding response values of various machining processes.

b) Combining certain DM tools into hybrid models can lead to their applications with superior performance, overcoming limitations of the individual DM tools. Thus, hybridization of DM algorithms may lead to high quality solutions as compared to the existing algorithms.

c) Optimization of the NN architecture using different metaheuristic algorithms may be researched.

d) Applications of advanced deep learning techniques for analyzing and obtaining the optimal parametric combinations of other NTM processes may be studied.

REFERENCES

7. REFERENCES

[1] Rajput RK. A Textbook of Manufacturing Technology: (Manufacturing Processes). Laxmi Publications Private Limited; India, 2008.

[2] Groover MP. Fundamental of Modern Manufacturing: Materials, Processes and Systems. John Wiley & Sons; U.S.A., 2020.

[3] Singh R. Introduction to Basic Manufacturing Processes and Workshop Technology. New Age Publication Limited; India, 2006.

[4] El-Hofy H. Fundamentals of Machining Processes: Conventional and Non Conventional Machining Processes. CRC Press; Boca Raton, 2007.

[5] Jain VK. Advanced Machining Processes. Allied Publishers Private Limited; India, 1980.

[6] Bhattacharyya B, Doloi B. Modern Machining Technology: Advanced, Hybrid, Micro machining and Super Finishing Technology. Academic Press; U.S.A., 2019.

[7] Gupta K, Gupta MK. Developments in nonconventional machining for sustainable production: A state-of-the-art review. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 2019; 233(12):4213-4232.

[8] Ho KH, Newman ST. State of the art electrical discharge machining (EDM). International Journal of Machine Tools and Manufacture 2003; 43(13):1287-1300.

[9] Puri AB, Bhattacharyya B. Modeling and analysis of white layer depth in a wire-cut EDM process through response surface methodology. The International Journal of Advanced Manufacturing Technology 2005; 25(3-4):301-307.

[10] Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining. American Association for Artificial Intelligence; U.S.A.,1996.

[11] Lee MH. Knowledge based factory. Artificial Intelligence in Engineering 1993; 8(2):109-125.

[12] Irani KB, Cheng J, Fayyad UM, Quian Z. Applying machine learning to semiconductor manufacturing. IEEE Expert 1993; 8(1):41-47.

[13] Piatetsky-Shapiro G. The data mining industry coming of age. IEEE Intelligence System and their Applications 1999;14(6):32-34.

[14] Oh S, Han J, Cho H. Intelligent process control system for quality improvement by data mining in the process industry. In: Data Mining for Design and Manufacturing: Methods and Applications. Kluwer Academic Publishers; U.S.A., 2001; 289-310.

[15] Skormin VA, Gorodetski VI, Popyack IJ. Data mining technology for failure of prognostic of avionics. IEEE Transactions on Aerospace and Electronic Systems 2002; 38(2):388-403.

[16] Chen WC, Tseng SS, Wang CY. A novel manufacturing defect detection method using association rule mining techniques. Expert Systems with Applications 2005; 29 (4):807-815.

[17] Shen L, Tay FEH, Qu LS, Shen Y. Fault diagnosis using rough set theory. Computers in Industry 2000; 43(1):61-72.

[18] Batanov D, Nagarur N, Nitikhunkasem P. EXPERT-MM: A knowledge-based system for maintenance management. Artificial Intelligence in Engineering 1993; 8(4):283-291.

[19] Romanowski CJ, Nagi R. Analyzing maintenance data using data mining methods. In: D. Braha (ed.), Data Mining for Design and Manufacturing: Methods and Applications, Kluwer Academic Publishers; U.S.A., 2001, 3:235-254.

[20] Harding JA, Shahbaz M, Srnivas, KusiakA. Data mining in manufacturing: A review, Journal of Manufacturing Science and Engineering 2006; 128(4):969-976.

[21] Rohanizadeh SS, Moghadam MB. A proposed data mining methodology and its application to industrial procedures. Journal of Industrial Engineering 2009; 4:37-50.

[22] Fayyad U, Stolorz P. Data mining and KDD: Promise and challenges. Future Generation Computer Systems1997;13(2-3):99-115.

[23] Quinlan JR. Induction of decision trees. Machine Learning 1986;1:81-106.

[24] Koonce DA, Fang CH, Tsai SC. A data mining tool for learning from manufacturing systems. Computers & Industrial Engineering 1997; 33(1-2):27-30.

[25] Meng TK, Butler C. Solving multiple response optimisation problems using adaptive neural networks. The International Journal of Advanced Manufacturing Technology 1997;13:666-675.

[26] Sluga A, Jermol M, Zupanič D, Mladenić D, 1998. Machine learning approach to machinability analysis. Computers in Industry 1998; 37(3):185-196.

[27] Maki H, Maeda, A, Morita T, Akimori H. Applying data mining to data analysis in manufacturing. In: Proceedings of the 7th International Conference on Advances in Production Management Systems, Germany, 1999; 324-331.

[28] Choudhary AK, Harding JA, Tiwari MK. Data mining in manufacturing: a review based on the kind of knowledge. Journal of Intelligent Manufacturing 2000; 20:501-521.

[29] Koonce DA, Tsai SC. Using data mining to find patterns in genetic algorithm solutions to a job shop schedule. Computers & Industrial Engineering 2000; 38(3):361-374.

[30] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 2001; 129(1):1-47.

[31] Dabbas RM, Chen HN. Mining semiconductor manufacturing data for productivity improvement-an integrated relational database approach. Computers in Industry 2001; 45(1):29-44.

[32] Baek JG, Kim CO, Kim SS. Online learning of the cause-and-effect knowledge of a manufacturing process. International Journal of Production Research 2002; 40(14):3275-3290.

[33] Kim IS, Son JS, Yarlagadda PK. A study on the quality improvement of robotic GMA welding process. Robotics and Computer-Integrated Manufacturing 2003;19(6):567-572.

[34] Li M, Feng S, Sethi IK, Luciow J, Wagner K. Mining production data with neural network & CART. In: Proceedings of the 3rd IEEE International Conference on Data Mining, U.S.A., 2003; 731-734.

[35] Last M, Kandel A. Discovering useful and understandable patterns in manufacturing data. Robotics and Autonomous Systems 2004; 49(3-4):137-152.

[36] Feng CXJ, Wang XF. Data mining techniques applied to predictive modelling of the knurling process. IIE Transactions 2004; 36(3):253-263.

[37] Agard B, Kusiak A. Data-mining-based methodology for the design of product families. International Journal of Production Research 2004; 42(15):2955-2969.

[38] Tseng TLB, Kwon Y, Ertekin YM. Feature-based rule induction in machining operation using rough set theory for quality assurance. Robotics and Computer-integrated Manufacturing 2005; 21(6):559-567.

[39] Chen WC, Tseng SS, Wang CY. A novel manufacturing defect detection method using association rule mining techniques. Expert Systems with Applications 2005; 29(4): 807-815.

[40] Pham DT, Afify AA. Machine-learning techniques and their applications in manufacturing. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 2005; 219(5):395-412.

[41] Jemwa GT, Aldrich C. Improving process operations using support vector machines and decision trees. American Institute of Chemical Engineers Journal 2005; 51(2):526-543.

[42] Öztürk A, Kayalıgil S, Özdemirel, NE. Manufacturing lead time estimation using data mining. European Journal of Operational Research 2006;173(2):683-700.

[43] Kusiak A. Data mining: manufacturing and service applications. International Journal of Production Research 2006; 44(18-19):4175-4191.

[44] Rokach L, Maimon O. Data mining for improving the quality of manufacturing: a feature set decomposition approach. Journal of Intelligent Manufacturing 2006; 17: 285-299.

[45] Harding JA, Shahbaz M, Srinivas, Kusiak A. Data mining in manufacturing: a review. Journal of Manufacturing Science and Engineering 2006;128:969-976.

[46] Feng JX, Kusiak A. Data mining applications in engineering design, manufacturing and logistics. International Journal of Production Research 2006; 44(14): 2689-2694.

[47] Da CC, Agard B, Kusiak A. Data mining for improvement of product quality. International Journal of Production Research, 2006; 44(18-19):4027-4041.

[48] Wang K. Applying data mining to manufacturing: the nature and implications. Journal of Intelligent Manufacturing 2007; 18:487-495.

[49] Wang K, Tong S, Eynard B, Roucoules L, Matta N. Application of data mining in manufacturing quality data. In: Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, China, 2007; 5382-5385.

[50] Cai-Yan L, You-Fa S. Application of data mining in production quality management. In: Proceedings of the Third International Symposium on Intelligent Information Technology Application, China, 2009; 284-287.

[51] Köksal G, Batmaz I, Testik MC. A review of data mining applications for quality improvement in manufacturing industry. Expert systems with Applications 2011; 38(10):13448-13467.

[52] Vazan P, Tanuska P, Kebisek M. The data mining usage in production system management. International Journal of Industrial and Manufacturing Engineering 2011; 5(5):922-926.

[53] Jha SK, Jha MVK, Prakash MP. Data mining application using decision tree and ANN for predicting surface roughness of end milling manufacturing process. International Journal of Computer Science Engineering and Information and Technology Research 2011;1(2):61-68.

[54] Ferreiro S, Sierra B. Comparison of machine learning algorithms for optimization and improvement of process quality in conventional metallic materials. The International Journal of Advanced Manufacturing Technology 2012; 60: 237-249.

[55] Corne D, Dhaenens C, Jourdan L. Synergies between operations research and data mining: The emerging use of multi-objective approaches. European Journal of Operational Research 2012; 221(3):469-479.

[56] Yacout S, Meshreki M, Attia H. Monitoring and control of machining process by data mining and pattern recognition. In: Proceedings of the Sixth International Conference on Complex, Intelligent, and Software Intensive Systems, Italy, 2012; 106-113.

[57] Bastos P, Lopes I, Pires L. A maintenance prediction system using data mining techniques. Proceedings of World Congress on Engineering 2012; 3:1448-1453.

[58] Gröger C, Niedermann F, Mitschang B. Data mining-driven manufacturing process optimization. Proceedings of the World Congress on Engineering 2012;3:4-6.

[59] Bastos P, Lopes I, Pires L. Application of data mining in a maintenance system for failure prediction. In: Safety, Reliability and Risk Analysis: Beyond the Horizon, 22nd European Safety and Reliability 2014; 1:933-940.

[60] Kurniadi KA, Ryu K, Kim D. Real-time parameter adjustment and fault detection of remote laser welding by using ANN. International Journal of Precision Engineering and Manufacturing 2014; 15:979-987.

[61] Accorsi R, Manzini R, Pascarella P, Patella M, Sassi S. Data mining and machine learning for condition-based maintenance. Procedia Manufacturing 2017;11:1153-1161.

[62] Ademujimi TT, Brundage MP, Prabhu VV. A review of current machine learning techniques used in manufacturing diagnosis. In: Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing: IFIP WG 5.7 International Conference, Hamburg, Germany, 2017; 407-415.

[63] Gurgenc T, Ucar F, Korkmaz D, Ozel C, Ortac Y. A study on the extreme learning machine based prediction of machining times of the cycloidal gears in CNC milling machines. Production Engineering 2019;13(6):635-647.

[64] Azadi M, Kolahan F. Modeling and optimization of the electrical discharge machining process based on a combined artificial neural network and particle swarm optimization algorithm. Scientia Iranica 2020; 27(3):1206-1217.

[65] Zhu Z, Ferreira K, Anwer N, Mathieu L, Guo K, Qiao L. Convolutional neural network for geometric deviation prediction in additive manufacturing. Procedia CIRP 2020;91:534-539.

[66] Abdulateef OF. Surface roughness prediction in turning operation of aluminum alloy 6061 using artificial neural network (ANN). Journal of Mechanical Engineering Research and Developments 2020;43:360-366.

[67] Khawaja AUH, Jahanzaib M, Munawar M. Optimizing the performance of high-speed machining on 15CDV6 HSLA steel in terms of green manufacturing using response surface methodology and artificial neural network. International Journal of Precision Engineering and Manufacturing 2021; 22:1125-1145.

[68] Palanisamy D, Manikandan N, Binoj JS, Ramesh R, Narayana TS. Performance comparison of artificial neural network and multiple regression models for wire electrical discharge machining of haste alloy. Materials Today: Proceedings 2021; 39:524-532.

[69] Nguyen VH, Le TT, Truong HS, Le MV, Ngo VL, Nguyen AT, Nguyen HQ. Applying Bayesian optimization for machine learning models in predicting the surface roughness in single-point diamond turning polycarbonate. Mathematical Problems in Engineering 2021; 2021(1):6815802.

[70] Bakhtiyari AN, Wang Z, Wang L, Zheng H. A review on applications of artificial intelligence in modeling and optimization of laser beam machining. Optics & Laser Technology 2021;135:106721.

[71] Bhattacharya S, Das PP, Chatterjee P, Chakraborty S.Prediction of responses in a sustainable dry turning operation: A comparative analysis. Mathematical Problems in Engineering 2021; 2021:9967970.

[72] Özden G, Mata F, Öteyaka MÖ. Artificial neural network modeling for prediction of cutting forces in turning unreinforced and reinforced polyamide. Journal of Thermoplastic Composite Materials 2021; 34(3):353-363.

[73] Konda N, Verma R, Jayaganthan R. Machine Learning based predictions of fatigue crack growth rate of additively manufactured Ti6Al4V. Metals 2021; 12(1):50.

[74] Elango S, Natarajan E, Varadaraju K, Gnanamuthu A, Morris E, Durairaj R, Mohanraj K, Osman MA. Extreme gradient boosting regressor solution for defy in drilling of materials. Advances in Materials Science and Engineering 2022; 2022(1):8330144.

[75] Wang F, Elbadawi M, Tsilova SL, Gaisford S, Basit AW, Parhizkar M. Machine learning predicts electrospray particle size. Materials & Design 2022; 219:110735.

[76] Ganeshkumar S, Deepika T, Haldorai A. A supervised machine learning model for tool condition monitoring in smart manufacturing. Defence Science Journal 2022; 72(5):712-720.

[77] Cheng M, Jiao L, Yan P, Li S, Dai Z, Qiu T, Wang X. Prediction and evaluation of surface roughness with hybrid kernel extreme learning machine and monitored tool wear. Journal of Manufacturing Processes 2022; 84:1541-1556.

[78] Kumar S, Gopi T, Harikeerthana N, Gupta MK, Gaur V, Krolczyk GM, Wu, C. Machine learning techniques in additive manufacturing: a state of the art review on design, processes and production control. Journal of Intelligent Manufacturing 2023; 34(1):21-55.

[79] Theeda S, Jagdale SH, Ravichander BB, Kumar G. Optimization of process parameters in laser powder bed fusion of SS316L parts using artificial neural networks. Metals 2023; 13(5):842.

[80] Huang PB, Inderawati MMW, Rohmat R, Sukwadi R. The development of an ANN surface roughness prediction system of multiple materials in CNC turning. The International Journal of Advanced Manufacturing Technology 2023;125(3):1193-1211.

[81] Özkavak HV, Ince M, Bıçaklı EE. Prediction of mechanical properties of the 2024 aluminum alloy by using machine learning methods. Arabian Journal for Science and Engineering 2023; 48(3):2841-2850.

[82] Yang X, El-Fallah GM, Tao Q, Fu J, Leng C, Shepherd J, Dong, H. Dimensionality reduction for machine learning using statistical methods: A case study on predicting mechanical properties of steels. Materials Today Communications 2023;34:105162.

[83] Barrionuevo GO, Walczak M, Ramos-Grez J, Sánchez-Sánchez X. Microhardness and wear resistance in materials manufactured by laser powder bed fusion: Machine learning approach for property prediction. CIRP Journal of Manufacturing Science and Technology 2023;43:106-114.

[84] Chakraborty S, Chakraborty S. Applications of artificial neural networks in machining processes: a comprehensive review International Journal on Interactive Design and Manufacturing 2024; 18:1917-1948.

[85] Farooq MU, Kumar R, Khan A, Singh J, Anwar S, Verma A, Haber R. Sustainable machining of Inconel 718 using minimum quantity lubrication: Artificial intelligence-based process modeling. Heliyon 2024; 10:e34836.

[86] Du J, Wang Y, Zhou X, Ji Z. Prediction model and optimization of energy consumption, cutting force, and surface roughness during machine tool cutting process based on high-order response surface methodology. Journal of the Brazilian Society of Mechanical Sciences and Engineering 2024; 46(7):444.

[87] Truong TT, Airao J, Hojati F, Ilvig CF, Azarhoushang B, Karras P, Aghababaei R. Data-driven prediction of tool wear using Bayesian regularized artificial neural networks. Measurement 2024; 238:115303.

[88] Zhao M, Xue B, Li B, Zhu J, Song W. Ensemble learning with support vector machines algorithm for surface roughness prediction in longitudinal vibratory ultrasound-assisted grinding. Precision Engineering 2024; 88:382-400.

[89] Kulandaiyappan NK., Stanislaus Arputharaj B, Rajendran P, Raja V, Karuppasamy A. Development of statistical and soft computing regression models for WEDM machining of aluminium composites. International Journal on Interactive Design and Manufacturing 2024; 1-17.

[90] Yau HT, Kuo PH, Hong SW. Milling wear prediction using an artificial neural network model. Engineering Applications of Artificial Intelligence 2024; 135:108686.

[91] Song G, Zhang J, Ge Y, Zhu K, Liu J, Yu L, Sun J. A multi-target predictive model for predicting tool wear and surface roughness. Expert Systems with Applications 2024; 251:123779.

[92] Thapaliya S, Valilai OF, Wicaksono H. Power consumption and processing time estimation of CNC machines using explainable artificial intelligence (XAI). Procedia Computer Science 2024; 232:861-870.

[93] Babu MS. An in-process machined surface roughness classification using an ensemble learning algorithm based on extracted automated features from real-time surface images in milling process. International Journal on Interactive Design and Manufacturing 2024; 1-13.

[94] Tan PN, Steinbach M, Kumar V. Introduction to Data Mining Instructor's Solution Manual. Pearson Addison Wesley; U.S.A., 2006.

[95] Han J, Kamber M, Pei J. Data Mining Concepts and Techniques. Elsevier Inc.; U.S.A., 2012.

[96] Dunham MH. Data mining: Introductory and Advanced Topics. Pearson Education; India, 2006.

[97] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, U.S.A., 1993; 207-216.

[98] Agarwal S, Dandge SS, Chakraborty S. Development of association rules to study the parametric influences in non-traditional machining processes. Sādhanā 2019; 44(11):230.

[99] Westphal C, Blaxton T. Data Mining Solutions: methods and tools for solving real-world problems. John Wiley & Sons; U.S.A.,1998.

[100] Susanti Y, Zukhronah E, Pratiwi H, Respatiwulan, Sri Sulistijowati H. Analysis of chi-square automatic interaction detection (CHAID) and classification and regression tree (CRT) for classification of corn production. IOP Conference Series: Journal of Physics 2017; 909:1-8.

[101] Rokach L, Maimon O. Data Mining with Decision Trees: Theory and Applications. World Scientific Publishing Co.; U.S.A., 2014.

[102] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Tree. Chapman and Hall; New York, 1993.

[103] Kass GV. An exploratory technique for investigating large quantities of categorical data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 1980; 29(2):119-127.

[104] Pitombo CS, De Souza AD, Lindner A. Comparing decision tree algorithms to estimate intercity trip distribution. Transportation Research Part C: Emerging Technologies 2017; 77:16-32.

[105] Sadoyan H, Zakarian A, Mohanty P. Data mining algorithm for manufacturing process control. International Journal of Advanced Manufacturing Technology 2006; 28:342-350.

[106] Chakraborty S, Bhattacharya S. Application of XGBoost algorithm as a predictive tool in a CNC turning process. Reports in Mechanical Engineering 2021; 2(1):190-201.

[107] Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis 2002; 38(4):367-378.

[108] Friedman JH, Popescu BE. Importance sampled learning ensembles. Journal of Machine Learning Research 2003; 94305:1-32.

[109] Que Q, Belkin M. Back to the future: Radial basis function networks revisited. In: Proceedings of 19th International Conference on Artificial Intelligence and Statistics, Spain, 2016; 1375-1383.

[110] Cressie N. The origins of kriging. Mathematical Geology 1990; 22:239-252.

[111] Ozkavak HV, Sofu MM, Duman B, Bacak S. Estimating surface roughness for different EDM processing parameters on Inconel 718 using GEP and ANN. CIRP Journal of Manufacturing Science and Technology 2021; 33:306-314.

[112] Bakhtiyari AN, Wang Z, Wang L, Zheng H. A review on applications of artificial intelligence in modelling and optimization of laser beam machining. Optics & Laser Technology 2021;135:106721.

[113] Srivastava A, Sharma A, Gaur AS, Kumar R, Modi YK. Prediction of surface roughness for CNC turning of EN8 steel bar using artificial neural network model. Journal Européen des Systèmes Automatisés 2019; 52(2):185-188.

[114] Accorsi R, Manzini R, Pascarella P, Patella M, Sassi S. Data mining and machine learning for condition-based maintenance. Procedia Manufacturing 2017; 11:1153-1161.

[115] Sharma A, Mathew L, Chatterji S, Goyal D. Artificial intelligence-based fault diagnosis for condition monitoring of electric motors. International Journal of Pattern Recognition and Artificial Intelligence 2020; 34(13):2059043.

[116] Kant G, Sangwan KS. Predictive modelling and optimization of machining parameters to minimize surface roughness using artificial neural network coupled with genetic algorithm. Procedia CIRP 2015; 31:453-458.

[117] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into Imaging 2018; 9:611-629.

[118] Cui W, Zhang Y, Zhang X, Li L, Liou F. Metal additive manufacturing parts inspection using convolutional neural network. Applied Sciences 2020; 10(2):545.

[119] Haykin SS. Neural networks: A Comprehensive Foundation. Prentice Hall; U.S.A., 1999.

[120] Sharma P, Singh A. Era of deep neural networks: A review. In: Proceedings of 8th International Conference on Computing Communication and Networking Technologies, India, 2017; 1-5.

[121] Nasir V, Sassani F. A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. The International Journal of Advanced Manufacturing Technology 2021; 115(9):2683-2709.

[122] Gupta A, Singh H, Agrawal A. Taguchi-fuzzy multi output optimization (MOO) in high speed CNC turning of AISI P-20 tool steel. Expert Systems with Applications 2011; 38:6822-6828.

[123] Sadoyan H, Zakarian A, Mohanty P. Data mining algorithm for manufacturing process control. International Journal of Advanced Manufacturing Technology 2006; 28:342-350.

[124] Kumar A, Kumar V, Kumar J. Multi-response optimization of process parameters based on response surface methodology for pure titanium using WEDM process, International Journal of Advanced Manufacturing Technology 2013; 68(9-12):2645-2668.

[125] Sarker B, Chakraborty S. Structural equation modeling-based performance estimation and parametric analysis of wire electrical discharge machining processes. Sadhana 2021; 46:1-14.

[126] Agarwal S, Dandge S, Chakraborty S. Development of association rules to study the parametric influences in non-traditional machining processes. Sadhana 2019; 44:230.

[127] Kataria R, Kumar J, Pabla BS. Experimental investigation and optimization of machining characteristics in ultrasonic machining of WC-Co composite using GRA method. Materials and Manufacturing Processes 2016; 31(5):685-693.

[128] Soundhar A, Zubar HA, Sultan MTBHH, Kandasamy J. Dataset on optimization of EDM machining parameters by using central composite design. Data in Brief 2019; 23:103671.

[129] Chen K, Chen H, Liu L, Chen S. Prediction of weld bead geometry of MAG welding based on XGBoost algorithm. International Journal of Advanced Manufacturing Technology 2019; 101(9):2283-2295.

[130] Somasundaram M, Kumar JP. Multi response optimization of EDM process parameters for biodegradable AZ31 magnesium alloy using TOPSIS and grey relational analysis. Sādhanā 2022; 47(3):136.

[131] Nallathambi I, Ramar R, Pustokhin DA, Pustokhina IV, Sharma DK, Sengan S. Prediction of influencing atmospheric conditions for explosion avoidance in fireworks manufacturing Industry-A network approach. Environmental Pollution 2022; 304:119182.

[132] Hedayat A, Davilu H, Barfrosh A, Sepanloo K. Estimation of research reactor core parameters using cascade feed forward artificial neural networks, Progress in Nuclear Energy 2009;51:709-718.

PUBLICATIONS

Scientia Iranica E (2021) 28(6), 3653–3674

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

Decision tree-based parametric analysis of a CNC turning process

S.S. Dandge[a] and S. Chakraborty[b,*]

a. *Department of Mechanical Engineering, Government Polytechnic, Murtizapur, Maharashtra, India.*
b. *Department of Production Engineering, Jadavpur University, Kolkata, West Bengal, India.*

Received 7 May 2019; received in revised form 7 January 2020; accepted 18 February 2020

KEYWORDS
CNC turning process;
Decision tree;
CART;
CHAID;
Parameter;
Response.

Abstract. Computer Numerical Control (CNC) is a manufacturing concept where machine tools are automated to perform some predefined functions based on the instructions fed to them. CNC turning processes have found wide-ranging applications in modern-day manufacturing industries due to their capabilities to produce low-cost high-quality parts/components with very close dimensional tolerances. In order to exploit the fullest potential of a CNC turning process, its different input parameters should always be set to the optimal level for operation. In this paper, two classification tree algorithms, i.e., Classification And Regression Tree (CART) and CHi-squared Automatic Interaction Detection (CHAID) are applied to study the effects of various turning parameters on the responses and identify the best machining conditions for a CNC process. It is perceived that the obtained settings almost match with the observations of the earlier researchers. The CART algorithm outperforms CHAID with respect to higher overall classification accuracy and lower prediction risk.

1. Introduction

In manufacturing industries, machining is one of the major operations performed to remove unwanted material from the workpiece surface to attain the desired shape of the final product/component while fulfilling the customers[needs. Therefore, machining involves shaping the component by removing material. This can be achieved by using a tool whose material is harder than the component to be molded, which is removed by shear deformation in the form of chips [1]. Amongst various machining operations, turning is the most popularly adopted process. To meet the increasing requirements of low-cost and high-quality products, higher productivity, higher dimensional accuracy, and lower surface finish, Computer Numerical Control (CNC) technology is constantly replacing traditional turning operations. The precision and accuracy that can be achieved through the use of CNC turning operations cannot be achieved by traditional material removal processes [2]. The performance of any of the machining operations can usually be characterized by the combination of its various input parameters and outputs (responses). The input parameters of CNC turning operation mainly include feed rate, cutting speed, depth of cut, spindle speed, tool nose radius, machining time, type of the work material, cutting tool type, cutting fluid used etc. On the other hand, Material Removal Rate (MRR), Surface Roughness (SR), the amplitude of vibration, Tool Life (TL), Power

*. *Corresponding author.*
E-mail address: s_chakraborty00@yahoo.co.in (S. Chakraborty)

doi: 10.24200/sci.2020.53497.3269

Consumption (PC), Cutting Force (CF), and acoustic emission are the major process outputs. In CNC machines, these input parameters need to be more effectively controlled so as to achieve the target response values, minimize machining time and cost, minimize tool wear, impart more flexibility, generate complicated shapes, provide higher repeatability, attain high dimensional tolerance, reduce power consumption etc. It has been observed that there exist complex relationships between the input parameters and responses in CNC turning operations. Understanding these relationships and proposing parametric optimization techniques to improve the performance of CNC turning processes play a pivotal role in a manufacturing environment.

Optimization of a CNC turning operation has already been identified as a complex problem due to the involvement of several different and often contradictory objectives, like maximization of MRR and minimization of SR, maximization of turning efficiency, and minimization of power consumption etc. It is the primary objective of any optimization tool to identify the optimal values of various CNC turning parameters so as to achieve better process performance [3]. Usually, in manufacturing industries, the concerned machinists select the most suitable settings of different input parameters based on their knowledge and expertise. Sometimes, machining data handbooks have also been consulted to meet these requirements. However, the proposed machining parameters are far from their optimal values which could hinder the goal of achieving better performance. Due to the rapid development of CNC technology and the availability of large amounts of related data, it is now impossible to achieve the best machining performance only by deploying conventional optimization technologies. Therefore, the need for more advanced tools is ardently felt to fulfill the abovementioned objective. In this paper, the application of a data mining tool in the form of the development of decision trees is explored to determine the optimal parametric settings of different input parameters in a CNC turning operation. Two decision tree generation algorithms, i.e. Classification And Regression Tree (CART) and CHi-squared Automatic Interaction Detection (CHAID) are employed here to investigate the effects of the considered CNC turning parameters on the responses. The relative performance of both the algorithms is also compared with respect to solution accuracy and prediction risk.

The 'If-Then' decision rules generated based on the applications of CART and CHAID classification algorithms constitute a more powerful knowledge representation to understand the effects of different input parameters on the responses for the considered CNC turning process. When these rules are organized as a non-overlapping decision set, they become quite easy to interpret, even by a non-technical end-user. They follow a general structure, i.e. if the given CNC turning conditions are met, then certain response values can be attained or predicted. They are probably the most interpretable prediction models, semantically resembling the natural language and human thinking process. They also provide valuable information on how to make the final decision and why certain conditions are met. The rule generation process using the CART and CHAID algorithms has high speed and scalability and is almost robust against the presence of outliers in the input dataset. The decision rules usually generate sparse models, which do not contain many features and draw final conclusions based on only a few binary statements. They can be generated from large-scale datasets containing numerical and categorical information.

2. Literature survey

Gupta et al. [4] employed Taguchi method along with fuzzy modeling for multi-objective optimization of a CNC turning process while considering cutting speed, feed rate, depth of cut, tool nose radius, and cutting environment as the input parameters, and SR, tool life, CF and power consumption as the responses. Mukherjee et al. [5] introduced the application of the Taguchi method to identify the optimal operating levels of speed, feed, and depth of cut to maximize MRR during CNC turning of SAE 1020 material. Marko et al. [6] applied the Particle Swarm Optimization (PSO) technique to determine the optimal settings of three CNC cutting parameters, i.e. cutting speed, feed rate, and cutting depth for achieving the desired values of cutting force, SR and tool life. Saini and Pradhan [7] conducted CNC turning operation on aluminum alloy 8011 to investigate the effects of cutting speed, feed, and depth of cut on MRR and SR using an integrated Taguchi-fuzzy approach. While taking into account tool nose radius, cutting speed, feed rate, and depth of cut as the input parameters, Vasudevan et al. [8] combined grey theory and fuzzy technique with the Taguchi method to optimize SR, tangential CF, and MRR in CNC turning of GFRP/epoxy composite materials. Saini and Pradhan [9] studied the effects of three CNC turning parameters, i.e. spindle speed, feed rate, and depth of cut on SR properties of 316L stainless steel, EN24 alloy steel, and Ti6Al4V alloy materials. While considering cutting speed, spindle speed, feed rate, and depth of cut in a CNC machining operation, Aghdeab et al. [10] determined the minimum SR values using the Simulated Annealing (SA) technique. Camposeco-Negrete [11] identified feed rate and depth of cut as the two most important input parameters in rough turning operation of AISI 6061 T6 aluminum material to minimize energy consumption and SR and maximize MRR. Sarıkaya and Güllü [12] presented the application of Taguchi-based Grey Relational Analysis (GRA) to identify the optimal combination of the type of the cutting fluid, fluid flow rate, and cutting speed for having better values of flank wear, notch wear, and SR during machining Haynes 25 material under minimum quantity lubrication condition.

Asiltürk et al. [13] presented the combined application of Taguchi methodology and Response Surface Methodology (RSM) to study the effects of spindle rotational speed, feed rate, depth of cut, and tool tip radius on SR properties of Co28Cr6Mo alloy. Kumar et al. [14] identified the optimal parametric combination of cutting speed, feed rate, and depth of cut during CNC machining of AISI 1045 steel material to have the most preferred values of tool wear rate and MRR. During the CNC turning of AA7075 material, Maheswara Rao and Venkatasubbaiah [15] concluded that feed and cutting speed were the most significant parameters affecting SR of the machined components. Klancnik et al. [16] employed gravitational search algorithm and non-sorting genetic algorithm-II as the multi-objective optimization tools for a CNC turning process. Cutting speed, depth of cut, and feed rate were considered as the input parameters, and SR, CF, and tool life were the responses for the said process. Based on Taguchi experimental design plan, Bilga et al. [17] studied the effects of cutting speed, feed rate, depth of cut, and nose radius on some energy consumption responses, like energy efficiency, power factor, and active energy consumed during CNC rough turning operation of EN 353 alloy steel material. Based on Taguchi's L_9 experimental design plan, Kushwaha and Singh [18] studied the effects of cutting speed, feed rate, and depth of cut on SR and MRR during CNC turning of Inconel 625 material using coated carbide tool. Based on a developed model, Nataraj and Balasubramanian [19] determined the optimal settings of three CNC turning parameters, i.e. cutting speed, depth of cut, and feed rate in order to minimize SR, the intensity of vibration, and work-tool interface temperature. Nayak and Sodhi [20] applied the RSM technique to evaluate the relationship between three CNC turning parameters, i.e. depth of cut, feed rate, and cutting speed, and two responses, i.e. MRR and SR. The desirability function approach was later adopted to determine the optimal parametric settings for the considered process. During CNC turning of aluminum alloy, Sahoo et al. [21] integrated weighted principal component analysis with RSM technique to minimize SR and tool vibration while taking into account spindle speed, feed rate, and depth of cut as the input parameters. Based on the developed RSM-based equations, Mandal et al. [22] determined the optimal values of spindle speed, feed rate, and depth of cut for achieving favorable values of MRR, SR, and power in a CNC turning process. The corresponding Pareto fronts for the conflicting objectives were also proposed. Suresh et al. [23] determined the optimal settings of cutting speed, feed rate, tool nose radius, and depth of cut in a CNC turning operation for having minimum SR and maximum MRR values.

Akkuş [24] considered cutting speed, feed rate, and depth of cut as the process parameters in CNC machining of AISI 1040 steel material and identified their optimal settings for having minimum SR values. Bhanu Prakash et al. [25] treated spindle speed, feed rate, and depth of cut as the input parameters, and MRR and SR as the responses during CNC turning operation of $AlSi_7$ Mg material. Taguchi method and GRA technique were later adopted to optimize the said process. Gadekula et al. [26] determined the best parametric settings for spindle speed, feed rate, and depth of cut when using the Taguchi method to perform CNC turning operations on high-carbon and high-chromium steel workpiece materials. The MRR and SR were the responses for the considered process. Palanisamy and Senthil [27] proposed a combined application of the grey system and fuzzy logic approach to optimize cutting speed, feed rate, and depth of cut in a CNC turning process for achieving minimum values of SR and power consumption. Sahoo et al. [28] developed second-order RSM-based regression models to investigate the effects of three CNC turning parameters, i.e. spindle speed, feed rate, and depth of cut on two responses, i.e. SR and tool vibration. Weighted Aggregate Sum Product Assessment (WASPAS) method was later adopted for parametric optimization of the considered process. While performing CNC turning operation on aluminum alloy, Saravanakumar et al. [29] employed Taguchi method to determine the optimal settings of feed, speed and depth of cut to attain minimum values of SR and roundness error. Using RSM technique, Nataraj et al. [30] examined the impacts of feed rate, cutting speed and depth of cut on the work-tool interface zone temperature and SR during CNC turning operation of LM6 reinforced metal matrix composites. Vasudevan et al. [31] combined principal component analysis with GRA technique to explore the effects of feed rate, depth of cut, cutting speed and tool nose radius on MRR, CF and different SR parameters during machining of glass fibre reinforced polymer composites. Rao et al. [32] performed multi-objective optimization of MRR and SR during CNC turning of stainless steel 304 work materials. Cutting speed, feed rate and depth of cut were taken into account as the input parameters for the said process. Vijay Kumar et al. [33] endeavoured to investigate the effects of feed rate, depth of cut and spindle speed on MRR and SR while machining EN 19 stainless steel materials using a CNC turning centre. Taguchi's L_{18} orthogonal array was incorporated for conducting the experimental trials. Arun Vikram et al. [34] applied Taguchi method to explore the interactions between feed rate, depth of cut and spindle speed, and SR, MRR and tool temperature. Later, GRA technique was employed to determine the best parametric condition for attaining better response values during wet machining of materials with low machinability. Chau et al. [35] integrated Taguchi method, adaptive neuro-fuzzy inference system and teaching-learning-based optimization algorithm for parametric optimization of a CNC turning process.

The existing literature is full of applications of various mathematical tools and techniques for parametric optimization of CNC turning processes. A great deal of the past researches is dedicated to the applications of the Taguchi method to identify the best settings of different turning parameters for attaining the desired response values. Various multi-criteria decision-making methods, like GRA, and WASPAS, have also been proposed for the same purpose. Based on the experimental data, some researchers have also attempted to develop the corresponding higher-order RSM-based regression equations to describe the relationships between various CNC turning parameters and responses. Metaheuristic algorithms, mainly in the form of SA, PSO, etc., have been later applied to solve those equations in order to identify the optimal values of various input parameters for enhancing the process performance. But, it is quite interesting to notice that there is an immense scarcity of the application of any kind of data mining tool for parametric optimization of CNC machining processes. In this paper, the application of a data mining tool in the form of decision (classification) trees is proposed for the first time for parametric analysis and optimization of a CNC turning process. Using CART and CHAID-based decision trees, the corresponding decision rules in the form of simple and understandable 'If-Then' statements are generated to study the effects of each of the considered CNC turning parameters on the responses. The relative performance of both these algorithms is also compared with respect to classification accuracy and prediction risk.

3. Decision tree

Over the years, decision trees have become one of the most popular tools in knowledge discovery and data mining. It mainly deals with exploring large amounts of data to find meaningful patterns [36]. Through applicating the decision tree algorithm the effort takes place to solve a given problem in the form of tree representation. It basically employs development of a training model which can predict class or value of target variables based on the decision rules generated from the initial dataset. It belongs to the class of supervised machine learning algorithm of data mining which has the capability to solve both regression and classification problems. It is also a non-parametric approach having no idea about the pertaining distribution of the data. The developed decision trees usually follow the human thinking process and the logic of interpreting the data is very strong. In a decision tree, the original dataset is broken down into smaller subsets while incrementally developing the associated decision tree at the same time. The final decision tree consists of many decision nodes and leaf nodes. A decision node may comprise two or more branches and a leaf node denotes a classification or a final decision. The topmost decision node containing the complete dataset is known as the root node. This dataset is sequentially split resulting in child nodes during classification. When no further splitting is possible, the final nodes are termed as terminal nodes. Similarly, each decision node of a decision tree relates to an attribute and each leaf node corresponds to a class label. In a decision tree, a series of 'If-Then' statements can be finally developed when tracing the path through the different decision and leaf nodes, starting from the root node. The decision trees have several advantages, likeability to deal with both categorical and numerical data, self-explanatory and being easy to interpret, scalability with big data, capability to process datasets having errors or missing values, the requirement of less computational effort, high predictive accuracy etc.

In this paper, two decision tree algorithms, i.e. CART and CHAID are applied for parametric analysis of a CNC turning process. In both the algorithms, each (non-terminal) node identifies a split condition which yields the optimal classification of dependent variables [37]. The details of these algorithms are presented here-in-under.

3.1. CART algorithm

It is a recursive partitioning method used for both regression and classification purposes. In this algorithm, the decision tree is constructed by splitting subsets of the data using all the predictor variables to create two child nodes repeatedly, beginning with the initial dataset. The best predictor variable is chosen based on a variety of impurity or diversity measures. The goal is to generate subsets of the data which are as homogeneous as possible with respect to the target variable [38]. The procedural steps for this algorithm are presented as below [39]:

Step 1: The algorithm runs through the entire dataset D to initiate the classification;

Step 2: If all datasets in D belong to class P, generate a node P and stop, otherwise, decide a predictor F and produce a decision node;

Step 3: Split the samples in D into possible subsets using a predictor selection measure called 'Gini index' which is used in splitting for classification to reduce the impurity of a node. Split at each node occurs only when it can generate the greatest improvement in classification accuracy;

Step 4: When selecting predictor variables, determine the best breakpoint so that the dependent variable can be best divided into two categories, which are characterized by the maximum internal uniqueness and external difference;

Step 5: For each split, the predictor variable with the best score of improvement is selected;

Step 6: Apply the algorithm recursively for all the subsets in D until all items or samples in a node have the same class, i.e. split is no longer possible;

Step 7: The process repeats recursively until one of the stopping rules is fulfilled:

a) If a node becomes pure, i.e. all cases in the node have identical values of the dependent variable;

b) If the current tree depth reaches the user-specified maximum limit;

c) If the size of a node is less than the user-specified minimum size;

d) If the split of a node results in a child node whose node size is less than the user-specified minimum size.

Step 8: Framing of 'If-Then' decision rules, i.e. rule: (condition) $\rightarrow Y$, where the condition is a combination of predictor variables and Y is the class label (decision).

In this algorithm, the measure of the importance of independent variables (X) in relation to a decision tree is defined as the sum of improvements that X has across all the splits in the tree when it is used as a primary or surrogate splitter. The importance of X is expressed in terms of a normalized quantity relative to the variable having the largest measure of importance. It ranges from 0 to 100, with the variable having the maximum importance score of 100. Thus, the variable importance plot is a good indicator that measures the importance of independent variables (which have already appeared in the decision tree).

3.2. CHAID algorithm

This algorithm, developed by Kass [40], is a decision tree development approach, based on the Chi-squared test, generated by repeatedly splitting the subsets into two or more child nodes starting with the initial dataset. Particularly, the predictor having the strongest relationship with the dependent variable based on p-value is utilized as the split node. To determine the best split at a particular node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. It is an exploratory data analysis method used to study the strongest association between a dependent variable and a large series of possible predictor variables which themselves may interact. The dependency measure may be a qualitative (nominal or ordinal) or a quantitative indicator. For qualitative (categorical) variables, a series of Chi-squared analyses is conducted between the dependent and predictor variables. For quantitative variables (continuous), F-test is used, where intervals (splits) are optimally determined for the predictor variables so as to maximize the ability to explain a dependent measure with respect to variance components [41]. This algorithm uses the following steps [40,41]:

Step 1: The first step is to create categorical predictor variables from continuous variables by dividing the respective continuous distributions into a given number of categories;

Step 2: In the merging stage, for each dependent variable, merge non-significant categories. It determines the pair of categories that is least significantly different (i.e., most similar) with respect to dependent variables. The most similar pair is the pair whose test statistic provides the largest p-value with respect to the dependent variable;

Step 3: If the statistical test for the given pair of predictor categories is not statistically significant, it will merge the respective predictor categories and find the next pair of categories which may now include the previously merged categories;

Step 4: If the statistical test for the given pair of predictor categories is statistically significant, the adjusted p-value is computed for the merged categories by applying the Bonferroni adjustments;

Step 5: In the splitting stage, the independent or predictor variable with the lowest significant p-value (calculated above) is selected as the best and the group is split on this predictor (i.e., each of the optimally merged categories of the predictor is used to define a subdivision of the parent group into a new subgroup). If no predictor has a significant p-value, the group is not split;

Step 6: The above-mentioned steps are repeated until all subgroups have either been analyzed or contain too few observations or cases.

The stopping rules are basically the same as described in the CART algorithm.

4. Decision trees for a CNC turning process

Based on Taguchi orthogonal array design plan, Gupta et al. [4] conducted 27 experiments considering cutting speed, feed rate, depth of cut, tool nose radius, and machining environment as the controllable parameters. On the other hand, Tool Life (TL) (in min), Power Consumption (PC) (in W), SR (in μm), and Cutting Force (CF) (in N) were the responses. It is worthwhile to mention here that among the considered responses, TL is the only 'Larger-The-Better' (LTB) type of quality characteristic, while, the remaining three are 'Smaller-The-Better' (STB) types. During experimentation, each of the CNC turning parameters was set at three different operating levels, as shown in Table 1. A high-speed CNC machining centre was utilized for conducting the experiments and AISI P20 tool steel bars (having a diameter of 65 mm and length 275 mm) were chosen as the work material. The results of the experimental study are provided in Table 2. In this table, the minimum, maximum, and median values for each of the responses are also shown.

For parametric analysis of the considered CNC turning process and investigating the effects of various input parameters on the process outputs (responses), the corresponding decision trees are developed using CART and CHAID algorithms in SPSS 16.0 software. For arriving at the best possible solutions, various parameters of the adopted decision tree algorithms are fine-tuned as follows.

For CART algorithm

- Growing method: CART;
- Categorical dependent variables: TL, PC, SR, and CF;
- Categorical independent variables: CS, FR, DOC, NR, and E;
- Validation: Cross-validation;
- Number of sample folds: 3;
- Growth limit: Maximum tree depth = 5;
- Minimum number of cases: Parent node = 3, Child node = 2;
- Impurity measure: Gini;
- Minimum change in improvement: 0.0001.

For CHAID algorithm

- Growing method: CHAID;
- Categorical dependent variables: TL, PC, SR and CF;
- Categorical independent variables: CS, FR, DOC, NR and E;
- Validation: Cross validation;
- Number of sample folds: 3;
- Growth limit: Maximum tree depth = 5;
- Minimum number of cases: Parent node = 3, Child node = 2.

Significance level for

a) Splitting node = 0.03;
b) Merging categories = 0.05;
c) Chi-square statistic = Pearson;

- Model estimation:

a) Maximum number of iterations = 100,
b) Minimum change in expected cell frequencies = 0.001;
c) Adjust significance values using the Bonferroni method.

Figure 1 exhibits the decision tree in the form of a classification tree diagram developed using the CART algorithm for tool life. In this diagram, tool life is represented as a dependent variable in the root node. As tool life is a continuous variable, its median value (27.66 min) calculated from the experimental dataset of Table 2 is adopted here for the splitting purpose. For tool life, which is an LTB quality characteristic, its values lower than or equal to 27.66 min are termed

Table 1. Computer Numerical Control (CNC) turning parameters and their levels [4].

CNC parameter	Symbol	Unit	Level		
			Low	**Medium**	**High**
Cutting Speed	CS	m/min	120	160	200
Feed Rate	FR	mm/rev	0.10	0.12	0.14
Depth Of Cut	DOC	mm	0.20	0.35	0.50
Nose Radius	NR	mm	0.40	0.80	1.20
Environment	E	—	Dry	Wet	Cryogenic

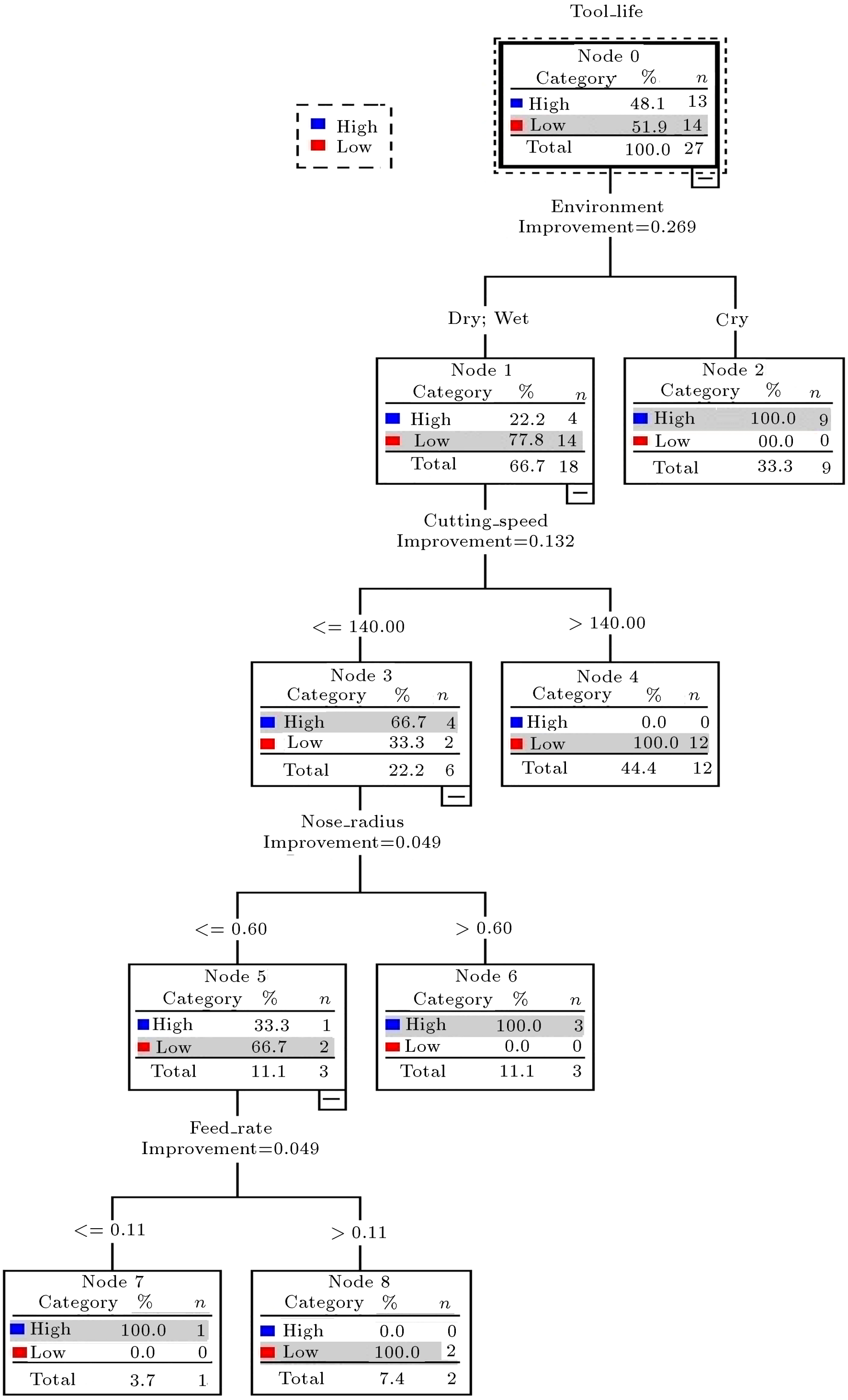


Figure 1. Classification tree for tool life using Classification And Regression Tree (CART) algorithm.

as 'low', whereas, values higher than 27.66 min are designated as 'high'. From the root node of the developed decision tree, it can be noticed that in the initial dataset, there are 13 experimental observations with high tool life and 14 observations have low tool life values. The first splitting is performed while taking the machining environment as the most important predictor variable. Between the two formed child nodes, node 2 appears to be a terminal node from where no further splitting can be possible. It is also identified as a pure node with no misclassification error. From node 1, taking cutting speed as the next important predictor variable, another classification is performed with the formation of node 4 as a terminal and pure

Table 2. Experimental data for the Computer Numerical Control (CNC) turning operation [4].

Exp. no.	CNC parameter					Response			
	CS	**FR**	**DOC**	**NR**	**E**	**TL**	**PC**	**SR**	**CF**
1	120	0.10	0.20	0.40	Dry	29.00	1066	1.41	171.30
2	120	0.10	0.35	0.80	Wet	34.00	1560	0.71	147.50
3	120	0.10	0.50	1.20	CRYO	54.67	866	0.60	111.74
4	120	0.12	0.20	0.80	Wet	34.67	1493	0.47	120.30
5	120	0.12	0.35	1.20	CRYO	51.66	987	0.19	180.60
6	120	0.12	0.50	0.40	Dry	27.00	1187	1.18	236.20
7	120	0.14	0.20	1.20	CRYO	50.00	960	0.67	157.70
8	120	0.14	0.35	0.40	Dry	24.66	1134	1.16	214.40
9	120	0.14	0.50	0.80	Wet	28.33	1813	0.92	286.90
10	160	0.10	0.20	1.20	Wet	27.66	1586	0.18	116.37
11	160	0.10	0.35	0.40	CRYO	47.66	1013	0.45	133.33
12	160	0.10	0.50	0.80	Dry	21.66	1240	0.43	191.23
13	160	0.12	0.20	0.40	CRYO	45.66	893	0.58	125.40
14	160	0.12	0.35	0.80	Dry	20.33	1253	0.72	149.43
15	160	0.12	0.50	1.20	Wet	25.66	1773	0.31	212.46
16	160	0.14	0.20	0.80	Dry	20.00	1107	0.66	162.93
17	160	0.14	0.35	1.20	Wet	22.33	1533	0.64	190.23
18	160	0.14	0.50	0.40	CRYO	41.33	1373	0.75	177.76
19	200	0.10	0.20	0.80	CRYO	40.00	1053	0.16	106.23
20	200	0.10	0.35	1.20	Dry	15.67	1373	0.23	208.50
21	200	0.10	0.50	0.40	Wet	21.67	2094	0.67	209.80
22	200	0.12	0.20	1.20	Dry	14.67	1286	0.40	200.20
23	200	0.12	0.35	0.40	Wet	20.33	1866	0.50	178.80
24	200	0.12	0.50	0.80	CRYO	37.66	1613	0.18	168.70
25	200	0.14	0.20	0.40	Wet	18.00	1573	0.64	162.00
26	200	0.14	0.35	0.80	CRYO	34.33	1453	0.31	162.00
27	200	0.14	0.50	1.20	Dry	16.66	1667	0.48	276.16
					Minimum	14.67	866	0.16	106.23
					Maximum	54.67	2094	1.41	286.90
					Median	27.66	1373	0.58	171.30

node. Two nodes, i.e. 5 and 6 now emerge out from node 3 using tool nose radius as the predictor variable. Finally, based on feed rate, the last two terminal nodes are constructed. This entire classification process along with the related characteristics is provided in Table 3. The percentages of correct classification at all the nodes are presented in Table 4. It can be observed that for tool life, there is no misclassification error in the decision tree developed using the CART algorithm. When the decision tree of Figure 1 and classification characteristics are analyzed in detail, several decision rules in the form of 'If-Then' statements are generated.

CART-based rules for tool life

Rule 1: If environment = cryogenic Then TL is

Table 3. Classification of tool life based on Classification And Regression Tree (CART).

Classification	Node	Characteristics
First	2	Cryogenic environment provides higher tool life
Second	1, 4	Dry or wet environment and $CS > 140$ m/min provide lower tool life
Third	1, 3, 6	Dry or wet environment, $CS \leq 140$ m/min and $NR > 0.60$ mm are responsible for higher tool life
Fourth	1, 3, 5, 7	Dry or wet environment, $CS \leq 140$ m/min, $NR \leq 0.60$ mm and $FR \leq 0.11$ mm/rev lead to higher tool life
Fifth	1, 3, 5, 8	Dry or wet environment, $CS \leq 140$ m/min, $NR \leq 0.60$ mm and $FR > 0.11$ mm/rev provide lower tool life

Table 4. Percentages of correct classification of tool life based on Classification And Regression Tree (CART).

Classification	Tool life			
	Low ($\leq$ 27.66 min)		High ($>$ 27.66 min)	
	Number of observations	Percentage	Number of observations	Percentage
1	0	0%	9	100%
2	12	100%	0	0%
3	0	0%	3	100%
4	0	0%	1	100%
5	2	100%	0	0%

(27.66–54.69]:

$$[P = 100\%, Q = 69.23\%, C = 33.33\%, QTY = 9]$$

$$[T = 202.56\%].$$

Rule 2: If environment = dry or wet and $CS >$ 140 m/min Then TL is [14.67–27.66]:

$$[P = 100\%, Q = 85.71\%, C = 44.44\%, QTY = 12]$$

$$[T = 230.15\%].$$

Rule 3: If environment = dry or wet, $CS \leq$ 140 m/min and $NR > 0.60$ mm, Then TL is (27.66–54.69]:

$$[P = 100\%, Q = 23.09\%, C = 11.11\%, QTY = 3]$$

$$[T = 134.20\%].$$

Rule 4: If environment = dry or wet, $CS \leq$ 140 m/min, $NR \leq 0.60$ mm and $FR \leq 0.11$ mm/rev Then TL is (27.66–54.69]:

$$[P = 100\%, Q = 7.70\%, C = 3.70\%, QTY = 1]$$

$$[T = 111.40\%].$$

Rule 5: If environment = dry or wet, $CS \leq$ 140 m/min, $NR \leq 0.60$ mm and $FR > 0.11$ mm/rev Then TL is [14.67–27.66]:

$$[P = 100\%, Q = 14.29\%, C = 7.41\%, QTY = 2]$$

$$[T = 121.70\%],$$

where P is the percentage of objects in the condition attribute set that corresponds to a rule (a measure of rule confidence), Q is the percentage of objects in the decision attribute set that corresponds to a rule, C is the percentage of objects that correspond to a rule (a measure of rule support) and QTY is the number of objects satisfying a particular rule. In this algorithm, $T(T = P+Q+C)$ represents the total strength (relative importance) of a rule [42].

Among these decision rules, Rule 2, having the maximum total strength of 230.15, states that when the said CNC turning operation is performed under a dry or wet environment and the cutting speed is greater than 140 m/min, the corresponding tool life would be low. On the other hand, Rule 1 with a total strength of 202.56 depicts that higher tool life can only be achievable under a cryogenic (CRYO) machining environment. It can also be noticed from these rules that low feed rate and low tool nose radius lead to higher tool life. The importance plot for tool life, as depicted in Figure 2, identifies the machining environment as the

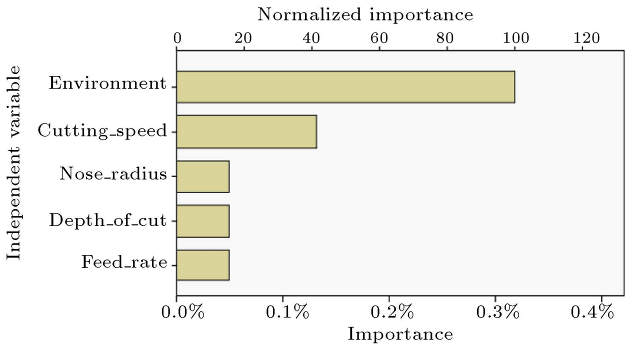


Figure 2. Importance plot for tool life .

most important CNC turning parameter, followed by cutting speed. Nose radius, depth of cut, and feed rate are observed to have the least importance on tool life. Similarly, the related decision tree for tool life is also generated using the CHAID algorithm, as exhibited in Figure 3. The classification characteristics and percentages of accurate classification at the identified nodes are provided in Tables 5 and 6 respectively. As compared to five classifications in the CART algorithm for tool life, there are only four classifications in the CHAID algorithm. Here, at the third classification in node 5, there is a 33.33% misclassification error. The corresponding rules generated using this algorithm are quite similar to those as developed by the CART algorithm. Machining environment and cutting speed are observed to be the two most important CNC turning parameters affecting tool life. While analyzing both these sets of decision rules, it can be concluded that for attaining higher tool life, cryogenic environment, and low values of cutting speed, tool nose radius and feed rate are always preferred. It is interestingly revealed that depth of cut appears to be an insignificant parameter having no effect on tool life. Based on analysis of variance (ANOVA) results for tool life, Gupta et al. [4] also identified machining

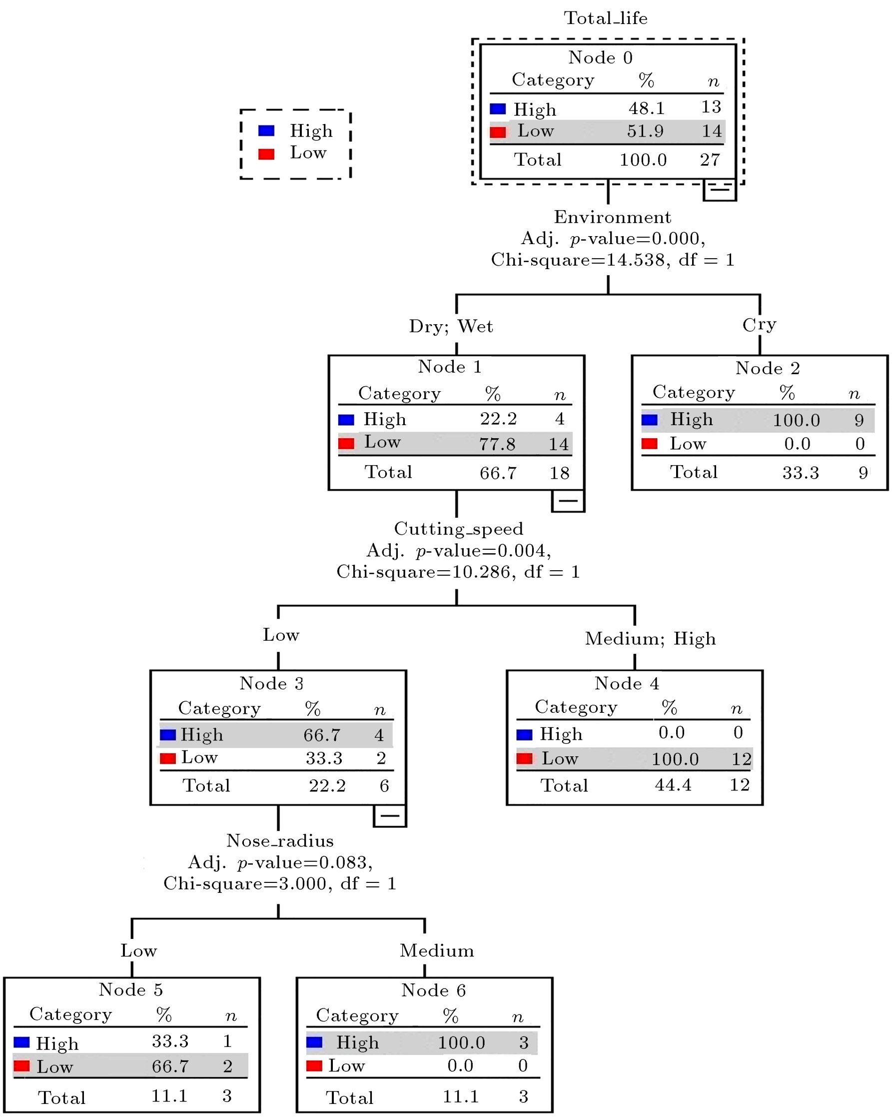


Figure 3. Classification tree for tool life using Chi-squared Automatic Interaction Detection (CHAID) algorithm.

Table 5. Classification of tool life based on Chi-squared Automatic Interaction Detection (CHAID).

Classification	Node	Characteristics
First	2	Cryogenic environment provides higher tool life.
Second	1,4	Dry or wet environment and medium or high CS (160 m/min or 200 m/min) lead to lower tool life.
Third	1,3,5	Dry or wet environment, low CS (120 m/min) and low NR (0.40 mm) are responsible for lower tool life.
Fourth	1,3,6	Dry or wet environment, low CS (120 m/min) and medium NR (0.80 mm) provide higher tool life.

Table 6. Percentages of accurate classification of tool life based on Chi-squared Automatic Interaction Detection (CHAID).

Classification	Tool life			
	Low ($\leq$ 27.66 min)		High ($>$ 27.66 min)	
	Number of observations	Percentage	Number of observations	Percentage
1	0	0%	9	100%
2	12	100%	0	0%
3	2	66.7%	1	33.3%
4	0	0%	3	100%

environment as the most important CNC turning parameter (69.27% contribution), followed by cutting speed (24.36% contribution). Depth of cut had almost no contribution (0.19%) on tool life. Using Signal-to-Noise (S/N) ratio values, it is recommended that the optimal combination of parameters for achieving a higher tool life are low cutting speed, low feed rate, low depth of cut, medium nose radius, and cryogenic environment which almost matches the combination proposed by a decision tree.

CHAID-based rules for tool life

Rule 1: If environment = cryogenic Then TL is (27.66–54.69]:

$$[P = 100\%, Q = 69.23\%, C = 33.33\%, QTY = 9]$$

$$[T = 202.56\%].$$

Rule 2: If environment = dry or wet and CS = medium or high Then TL is [14.67–27.66]:

$$[P = 100\%, Q = 85.71\%, C = 44.44\%, QTY = 12]$$

$$[T = 230.15\%].$$

Rule 3: If environment = dry or wet, CS = low and NR = low, Then TL is (27.66–54.69]:

$$[P = 66.70\%, Q = 14.28\%, C = 7.40\%, QTY = 2]$$

$$[T = 88.38\%].$$

Rule 4: If environment = dry or wet, CS = low and NR = medium, Then TL is [14.67–27.66]:

$$[P = 100\%, Q = 23.07\%, C = 11.11\%, QTY = 3]$$

$$[T = 134.18\%].$$

The decision tree for power consumption, which is developed using the CART algorithm, is shown in Figure 4. When the power consumption is less than or equal to 1373 W, it is designated as 'low' and when it is greater than 1373 W, its value is 'high'. As it is an STB type of response, its 'low' values are always preferred. The 'If-Then' rules extracted from the decision tree of Figure 4 highlight that dry or cryogenic machining environment and cutting speed less than or equal to 180 m/min always lead to lower power consumption (Rule 2 with total strength 224.44). Thus, a wet environment is responsible for higher power consumption (Rule 1 with total strength 208.33). Low feed rate and low depth of cut cause lower power consumption. The important plot of Figure 5 identifies the machining environment as the most critical CNC turning parameter affecting power consumption, followed by cutting speed. Interestingly, tool nose radius plays no significant role

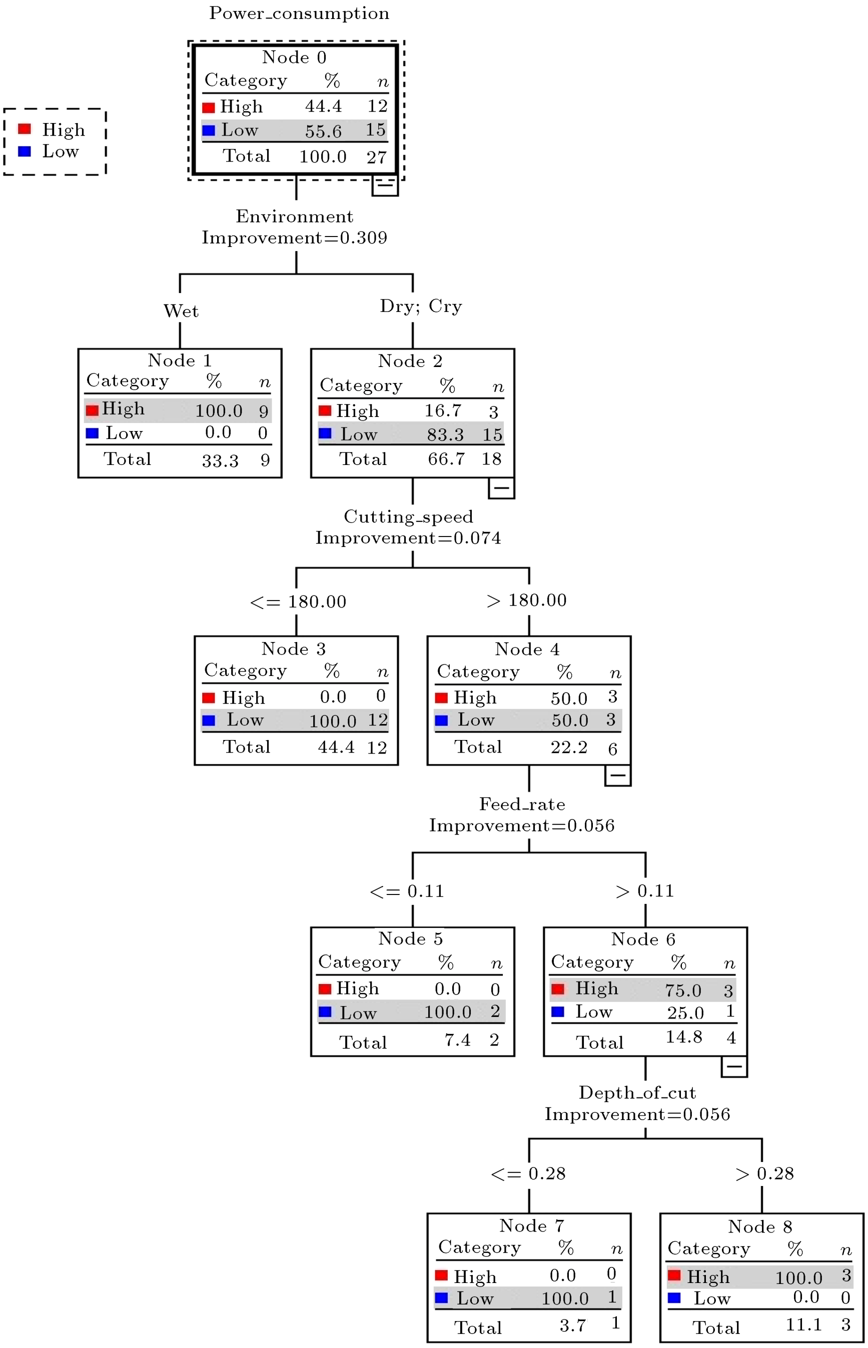


Figure 4. Classification tree for power consumption using Classification And Regression Tree (CART) algorithm.

in power consumption. The rules developed from the decision trees generated using CHAID algorithm (not shown here due to lack of space) also confirm these observations. Combining both the sets of decision rules from CART and CHAID algorithms, the optimal parametric mix of dry or cryogenic environment, low or medium cutting speed, low feed rate, and low nose radius would always lead to lower power consumption. Gupta et al. [4] also observed that machining environment = cryogenic, cutting speed = low, feed rate = low, depth of cut = low, and nose radius = medium were responsible for attaining the most desirable value of lower power consumption in the said CNC turning centre.

CART-based rules for power consumption

Rule 1: If environment = wet Then PC is (1373–2094]:

$$[P = 100\%, Q = 75\%, C = 33.33\%, QTY = 9]$$

$$[T = 208.33\%].$$

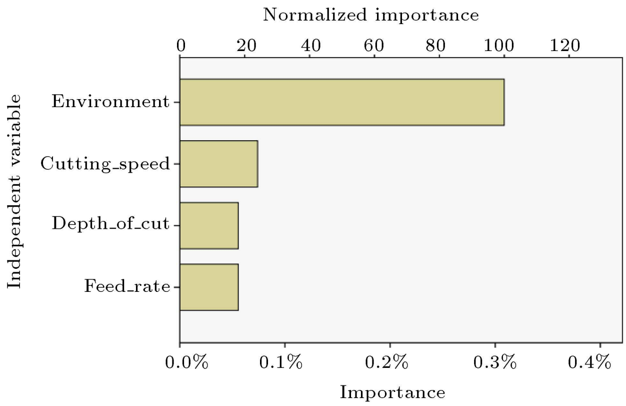


Figure 5. Importance of Computer Numerical Control (CNC) turning parameters affecting power consumption.

Rule 2: If environment = dry or cryogenic and $CS \leq$ 180 m/min Then PC is [866–1373]:

$[P = 100\%, Q = 80\%, C = 44.44\%, QTY = 12]$

$[T = 224.44\%]$.

Rule 3: If environment = dry or cryogenic, $CS >$ 180 m/min and $FR \leq 0.11$ mm/rev Then PC is [866–1373]:

$[P = 100\%, Q = 13.33\%, C = 7.40\%, QTY = 2]$

$[T = 120.73\%]$.

Rule 4: If environment = dry or cryogenic, CS > 180 m/min, $FR > 0.11$ mm/rev and $DOC \leq$ 0.28 mm Then PC is [866–1373]:

$[P = 100\%, Q = 6.67\%, C = 3.70\%, QTY = 1]$

$[T = 110.37\%]$.

Rule 5: If environment = dry or cryogenic, CS > 180 m/min, $FR > 0.11$ mm/rev and $DOC >$ 0.28 mm Then PC is [866–1373]:

$[P = 100\%, Q = 25\%, C = 11.11\%, QTY = 3]$

$[T = 136.11\%]$.

CHAID-based rules for power consumption

Rule 1: If environment = wet Then PC is (1373–2094]:

$[P = 100\%, Q = 75\%, C = 33.33\%, QTY = 9]$

$[T = 208.33\%]$.

Rule 2: If environment = dry or cryogenic and CS = low or medium Then PC is [866–1373]:

$[P = 100\%, Q = 80\%, C = 44.44\%, QTY = 12]$

$[T = 224.44\%]$.

Rule 3: If environment = dry or cryogenic, CS = high and FR = low Then PC is [866–1373]:

$[P = 100\%, Q = 13.33\%, C = 7.40\%, QTY = 2]$

$[T = 120.73\%]$.

Rule 4: If environment = dry or cryogenic, CS = high and FR = medium Then PC is [866–1373]:

$[P = 50\%, Q = 6.67\%, C = 3.70\%, QTY = 1]$

$[T = 60.37\%]$.

Rule 5: If environment = dry or cryogenic, CS = high and FR = high then PC is (1373–2094]:

$[P = 100\%, Q = 16.67\%, C = 7.40\%, QTY = 2]$

$[T = 124.07\%]$.

In Figure 6, the decision tree for SR which is developed using the CART algorithm is exhibited. The corresponding 'If-Then' rules are also subsequently generated. In this case, the SR values less than or equal to 0.58 μm are termed as 'low' (satisfactory) and those with greater than 0.58 μm values are styled as 'high'. Rule 1 with the maximum total strength of 148.40 reveals that when the machining environment is cryogenic and cutting speed is more than 140 m/min, the SR of the turned components would be satisfactory (low). A high nose radius also provides lower SR (Rule 3 with a total strength of 137.52). Similarly, a high feed rate is responsible for poor SR. The rules extracted from the decision tree which is developed using the CHAID algorithm (not shown here due to lack of space) prove that low or medium feed rate achieves better SR. In both the sets of rules, depth of cut appears to be an unimportant CNC turning parameter having no impact on SR. The importance of each of the turning parameters on SR is depicted in Figure 7 which clearly reveals the fact that machining environment and cutting speed are the two most significant parameters affecting SR. Depth of cut is the least important turning parameter. Gupta et al. [4] identified that a parametric combination of cutting speed = high, feed rate = medium, depth of cut = medium, nose radius = high, and machining environment = cryogenic would provide better SR of the turned components.

CART-based rules for SR

Rule 1: If environment = cryogenic and $CS >$ 140 m/min Then SR is [0.16–0.58]:

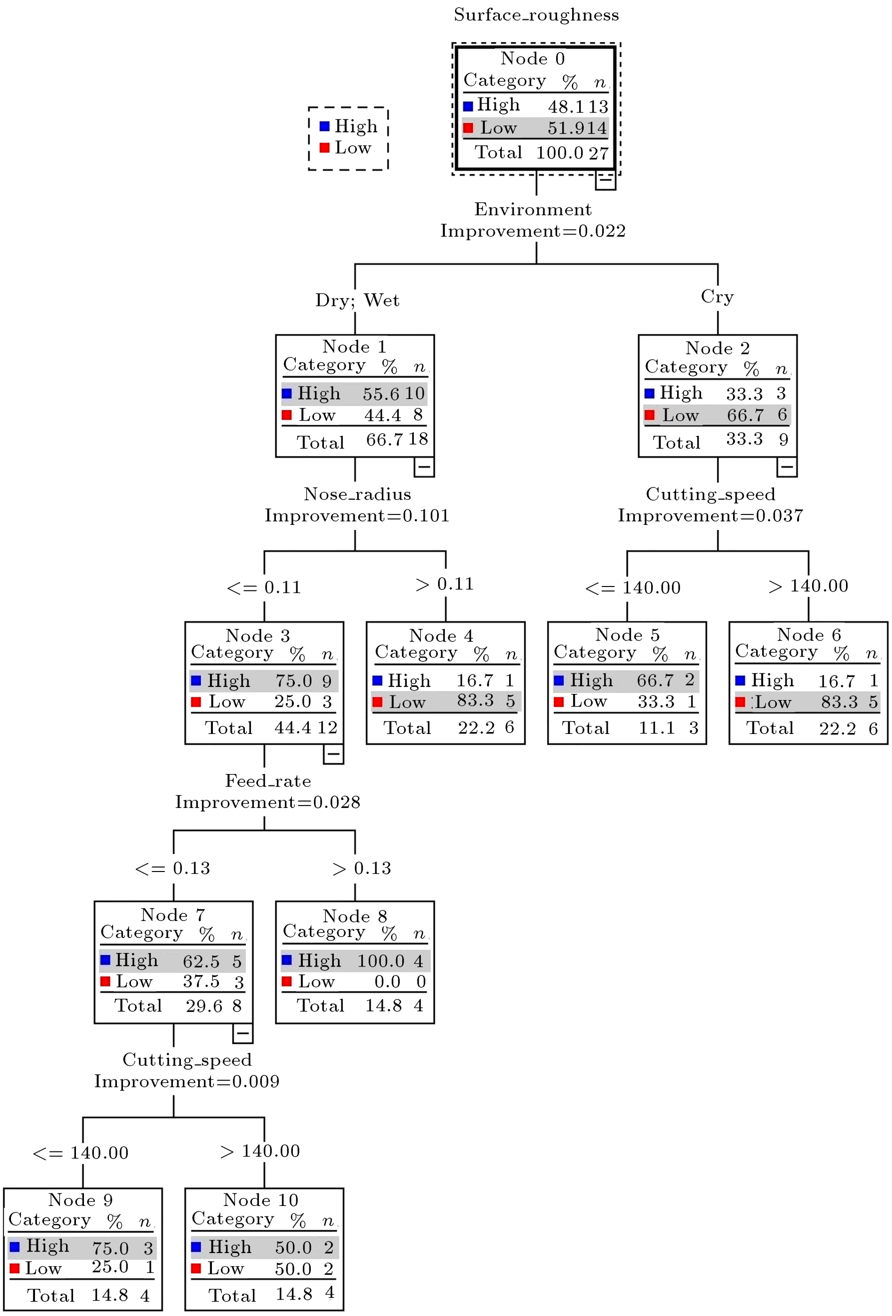


Figure 6. Classification tree for Surface Roughness (SR) using Classification And Regression Tree (CART) algorithm.

$$[P = 83.33\%, Q = 42.85\%, C = 22.22\%, QTY = 6]$$

$$[T = 148.40\%].$$

Rule 2: If environment = cryogenic and $CS \leq 140$ m/min Then SR is (0.58–1.41]:

$$[P = 66.70\%, Q = 15.38\%, C = 7.40\%, QTY = 2]$$

$$[T = 89.48\%].$$

Rule 3: If environment = dry or wet and $NR > 1.00$ mm Then SR is [0.16–0.58]:

$$[P = 83.30\%, Q = 35.71\%, C = 18.51\%, QTY = 5]$$

$$[T = 137.52\%].$$

Rule 4: If environment = dry or wet, $NR \leq 1.00$ mm and $FR > 0.13$ mm/rev Then SR is (0.58–1.41]:

$$[P = 100\%, Q = 30.77\%, C = 14.81\%, QTY = 4]$$

$$[T = 145.58\%].$$

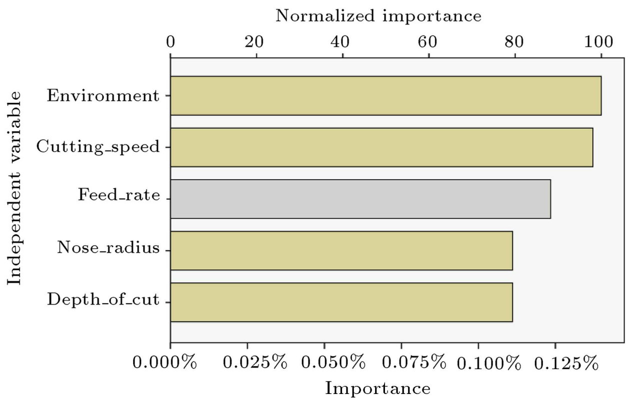


Figure 7. Importance plot for Surface Roughness (SR).

Rule 5: If environment = dry or wet, NR $\leq$ 1.00 mm, $FR \leq 0.13$ mm/rev and $CS \leq 140$ m/min Then SR is (0.58–1.41]:

$$[P = 75\%, Q = 23.07\%, C = 11.11\%, QTY = 3]$$

$$[T = 109.18\%].$$

Rule 6: If environment = dry or wet, $NR \leq$ 1.00 mm, $FR \leq 0.13$ mm/rev and $CS > 140$ m/min Then SR is (0.58–1.41]:

$$[P = 50\%, Q = 15.38\%, C = 7.41\%, QTY = 2]$$

$$[T = 72.79\%].$$

CHAID-based rules for SR

Rule 1: If environment = cryogenic and CS = high Then SR is [0.16–0.58]:

$$[P = 100\%, Q = 21.43\%, C = 11.11\%, QTY = 3]$$

$$[T = 132.54\%].$$

Rule 2: If environment = cryogenic, CS = low or medium and FR = high Then SR is (0.58–1.41]:

$$[P = 100\%, Q = 15.38\%, C = 7.41\%, QTY = 2]$$

$$[T = 122.79\%].$$

Rule 3: If environment = cryogenic, CS = low or medium and FR = low or medium Then SR is [0.16–0.58]:

$$[P = 75\%, Q = 21.43\%, C = 11.11\%, QTY = 3]$$

$$[T = 107.54\%].$$

Rule 4: If environment = dry or wet, NR = high and CS = high Then SR is [0.16–0.58]:

$$[P = 100\%, Q = 21.43\%, C = 11.11\%, QTY = 3]$$

$$[T = 132.54\%].$$

Rule 5: If environment = dry or wet, NR = high and CS = medium Then SR is [0.16–0.58]:

$$[P = 66.70\%, Q = 14.28\%, C = 7.41\%, QTY = 2]$$

$$[T = 88.39\%].$$

Rule 6: If environment = dry or wet, NR = low or medium and FR = high Then SR is (0.58–1.41]:

$$[P = 100\%, Q = 30.77\%, C = 14.81\%, QTY = 4]$$

$$[T = 145.58\%].$$

Rule 7: If environment = dry or wet, NR = low or medium and FR = low or medium Then SR is (0.58–1.41]:

$$[P = 62.50\%, Q = 38.46\%, C = 18.52\%, QTY = 5]$$

$$[T = 119.48\%].$$

The decision tree for cutting force originated from CART algorithm is shown in Figure 8. The corresponding 'If-Then' rules are subsequently generated from this decision tree. When the values of cutting force are less than or equal to 171.30 N, they are denoted as 'low' and when its values are greater than 171.30 N, they are designated as 'high'. An analysis of these rules reveals that when the machining environment is cryogenic, cutting speed is less than or equal to 180 m/min and feed rate is less than or equal to 0.11 mm/rev, and the achievable cutting force would be low. Similarly, a high depth of cut leads to higher cutting force. The rules extracted from the decision tree based on CHAID algorithm (not presented here) state that cryogenic environment would always provide lower cutting force. On the other hand, low depth of cut and low or medium nose radius are often responsible for attaining lower cutting force. When the relative importance of all the considered CNC turning parameters is plotted in Figure 9, it determines that the depth of cut is the most important parameter that affects the cutting force, followed by the machining environment and feed rate. Nose radius appears to be an insignificant CNC turning parameter for cutting force. An optimal parametric mix of moderate cutting speed, low feed rate, low depth of cut, moderate nose radius and cryogenic environment was identified by Gupta et al. [4] for lower cutting force, which almost corroborates with the decision trees-based observations.

CART-based rules for cutting force

Rule 1: If environment = cryogenic and $CS >$ 180 m/min Then CF is [106.23–171.30]:

$$[P = 100\%, Q = 21.43\%, C = 11.11\%, QTY = 3]$$

$$[T = 132.54\%].$$

Rule 2: If environment = cryogenic, $CS \leq$ 180 m/min and $FR \leq 0.11$ mm/rev Then CF is [106.23–171.30]:

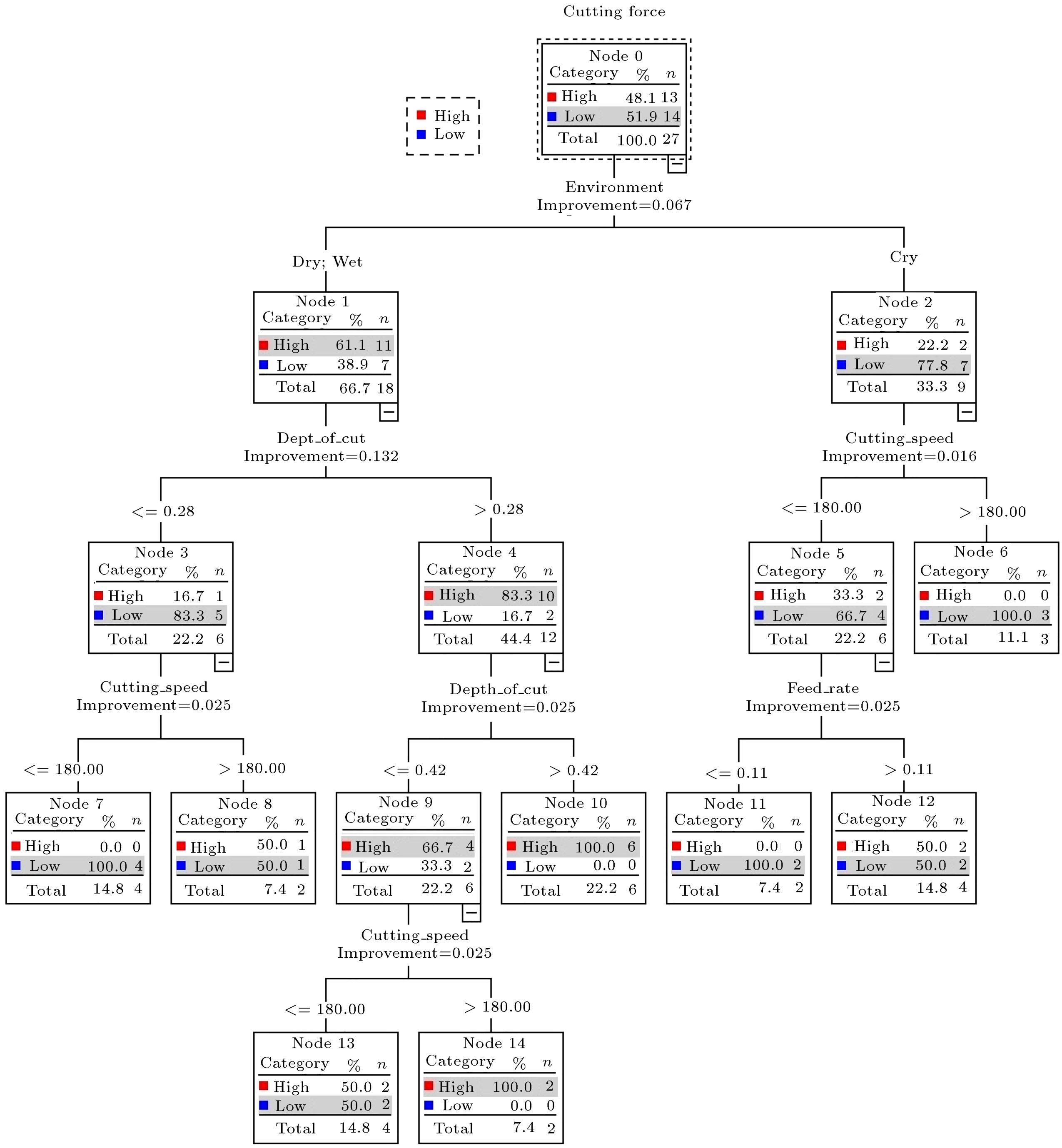


Figure 8. Classification tree for cutting force using Classification And Regression Tree (CART) algorithm.

$$[P = 100\%, Q = 14.28\%, C = 7.41\%, QTY = 2]$$

$$[T = 121.69\%].$$

Rule 3: If environment = cryogenic, $CS \leq 180$ m/min and $FR > 0.11$ mm/rev Then CF is [106.23–171.30]:

$$[P = 50\%, Q = 14.28\%, C = 7.41\%, QTY = 2]$$

$$[T = 71.69\%].$$

Rule 4: If environment = dry or wet, $DOC \leq 0.28$ mm and $CS \leq 180$ m/min Then CF is [106.23–171.30]:

$$[P = 100\%, Q = 28.57\%, C = 14.81\%, QTY = 4]$$

$$[T = 143.38\%].$$

Rule 5: If environment = dry or wet, $DOC \leq 0.28$ mm and $CS > 180$ m/min then CF is [106.23–171.30]:

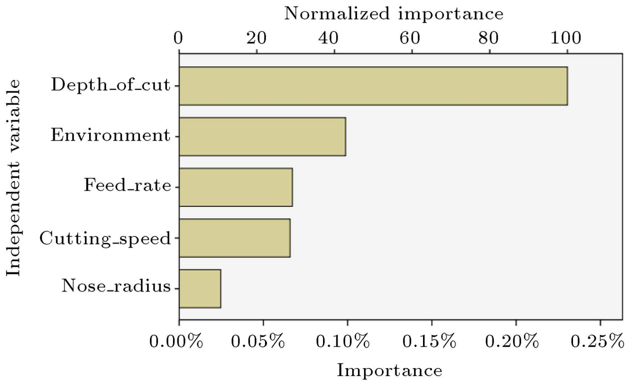


Figure 9. Importance of Computer Numerical Control (CNC) turning parameters affecting cutting force.

$[P = 100\%, Q = 7.69\%, C = 3.70\%, QTY = 1]$

$[T = 111.39\%]$.

Rule 6: If environment = dry or wet, $DOC >$ 0.28 mm Then CF is (171.30–286.90]:

$[P = 100\%, Q = 46.15\%, C = 22.22\%, QTY = 6]$

$[T = 168.37\%]$.

Rule 7: If environment = dry or wet, $DOC >$ 0.28 mm or $\leq$ 0.42 mm, and $CS \leq 180$ m/min Then CF is [106.23–171.30]:

$[P = 50\%, Q = 14.28\%, C = 7.41\%, QTY = 2]$

$[T = 71.69\%]$.

Rule 8: If environment = dry or wet, $DOC >$ 0.28 mm or $\leq$ 0.42 mm, and $CS > 180$ m/min Then CF is (171.30–286.90]:

$[P = 100\%, Q = 15.38\%, C = 7.41\%, QTY = 2]$

$[T = 122.79\%]$.

CHAID-based rules for cutting force

Rule 1: If environment = cryogenic Then CF is [106.23–171.30]:

$[P = 77.80\%, Q = 50\%, C = 25.92\%, QTY = 7]$

$[T = 153.72\%]$.

Rule 2: If environment = dry or wet and DOC = low Then CF is [106.23–171.30]:

$[P = 83.30\%, Q = 35.71\%, C = 18.52\%, QTY = 5]$

$[T = 136.53\%]$.

Rule 3: If environment = dry or wet, DOC = medium or high and NR = low or high Then CF is (171.30–286.90]:

$[P = 100\%, Q = 61.54\%, C = 29.63\%, QTY = 8]$

$[T = 191.17\%]$.

Rule 4: If environment = dry or wet, DOC = medium or high and NR = medium Then CF is [106.23–171.30]:

$[P = 50\%, Q = 14.29\%, C = 7.41\%, QTY = 2]$

$[T = 71.70\%]$.

In Table 7, a comparison of the classification accuracies for CART and CHAID algorithms for all the four responses is provided. From this table, it can be noted that for tool life and power consumption responses, CART algorithm can perfectly predict low and high tool life, and low and high power consumption values. The classification accuracies for high and low SR are 84.6% and 78.6% respectively. Similarly, using CART algorithm, high and low cutting forces can be predicted with accuracies of 85.7% and 76.9% respectively. Thus, CART algorithm can almost perfectly predict both the high and low values of all the considered responses, although it has a slightly greater tendency to accurately estimate high values of the responses. In case of CHAID algorithm, it can perfectly predict low values of tool life, power consumption and cutting force. High values of tool life and power consumption are predicted with 92.3% and 91.7% accuracies respectively. It has prediction accuracies of 84.6% and 71.4% for high and low SR values respectively. The classification accuracy for high cutting force is only 61.5%. Thus, it can be concluded that CHAID algorithm performs better in predicting low values of the considered responses. The overall classification accuracies of both these algorithms for the four responses are provided in Table 8. With respect to overall classification accuracy, CART algorithm outperforms CHAID in almost exactly predicting the responses of the CNC turning process under consideration. The corresponding values of Standard Error (SE) for CART are also comparatively low as compared to CHAID algorithm.

In order to visualize the effects of changing values of the responses of the considered CNC turning process on the prediction performance of CART and CHAID algorithms, a sensitivity analysis study is performed here. In this approach, incremental changes are made in the response values of the experimental dataset based on the equation:

$$R_N = R_O + (2 \times RAND() - 1) \times E \times R_O,$$

where R_O is the original response value, $RAND()$ is a uniform random number generator function between 0

Table 7. Classification accuracies for Tool Life (TL), Power Consumption (PC), Surface Roughness (SR), and Cutting Force (CF) using Classification And Regression Tree (CART) and Chi-squared Automatic Interaction Detection (CHAID) algorithms.

Response	Observed	CART			CHAID		
		Predicted				Predicted	
		High (> 27.66 min)	Low (≤ 27.66 min)	Percent correct	High (> 0.58 μm)	Low (≤ 0.58 μm)	Percent correct
TL	High (> 27.66 min)	13	0	100%	12	1	92.3%
	Low (≤ 27.66 min)	0	14	100%	0	14	100%
	Overall percentage	48.1%	51.9 %	100%	44.4%	55.6%	96.2%
		Predicted				Predicted	
	Observed	High (> 1373 W)	Low (≤ 1373 W)	Percent correct	High (> 1373 W)	Low (≤ 1373 W)	Percent correct
PC	High (> 1373 W)	12	0	100%	11	1	91.7%
	Low (≤ 1373W)	0	15	100%	0	15	100%
	Overall percentage	44.4%	55.6 %	100%	40.7%	59.3%	95.8%
		Predicted				Predicted	
	Observed	High (> 0.58 μm)	Low (≤ 0.58 μm)	Percent correct	High (> 0.58 μm)	Low (≤ 0.58 μm)	Percent correct
SR	High (> 0.58 μm)	11	2	84.6 %	11	2	84.6%
	Low (≤ 0.58 μm)	3	11	78.6 %	4	10	71.4%
	Overall percentage	51.8 %	48.1 %	81.6 %	55.5%	44.4%	78.0%
		Predicted				Predicted	
	Observed	High (> 171.30 N)	Low (≤ 171.30 N)	Percent correct	High (> 0.58 μm)	Low (≤ 0.58μm)	Percent correct
CF	High (> 171.30 N)	12	1	85.7 %	8	5	61.5%
	Low (≤ 171.30 N)	4	10	76.9 %	0	14	100%
	Overall percentage	59.2 %	40.8 %	81.3 %	29.6%	70.4%	80.7%

and 1, E is the relative error level and R_N is the new perturbed response value.

The relative error levels are set here as 5, 10, 15, 20, and 25%. The classification accuracies of both the algorithms at varying errors levels are provided in Table 9. It can be clearly propounded that the prediction performance of the CART algorithm is least affected by the changing response values in the experimental dataset and it is a more robust technique as compared to the CHAID algorithm.

Over the past few decades, decision tree algorithms, like CART and CHAID, have been in extensive use for solving predictive analytics problems. As they are generic models based on effective calculation procedures, they can easily arrive at the optimal solutions for a given classification/prediction problem. Decision trees generated by these algorithms are efficient managerial tools that present all the decisions/outcomes in the form of a flowchart with branches and leaves. Decision trees thus solve problems of machine learning by transforming the data into a tree representation. Each branch of the tree symbolizes a decision option.

Table 8. Risk of classifying Surface Roughness (SR), Tool Life (TL), Cutting Force (CF) and Power Consumption (PC).

Response	Method	Accuracy	Standard error
TL	CART	1.000	0.001
	CHAID	0.963	0.036
PC	CART	1.000	0.001
	CHAID	0.958	0.036
SR	CART	0.815	0.075
	CHAID	0.778	0.080
CF	CART	0.818	0.075
	CHAID	0.807	0.036

The leaves at the end of the branches show the possible outcomes. Decision trees can deal with quantitative, qualitative, or categorical attributes by assigning objects to a specific class in a classification problem. A decision tree is one of the simplest and most popular classification algorithms to learn, understand and interpret. They have several advantages, like the requirement of less computational effort for data preparation during pre-processing, no need for normalization and scaling of data, least affectability towards the missing observation in the dataset, provision of explanation about how a particular decision has been reached etc. Similarly, they also suffer from some disadvantages, like the requirement of higher time for training, a small change in data may cause a large change in the tree structure causing instability, inability to be applied for regression, and prediction of continuous variables.

5. Conclusions

This paper deals with the application of a data mining tool in the form of the development of decision trees using Classification And Regression Tree (CART) and Chi-squared Automatic Interaction Detection (CHAID) algorithms to determine the most preferable combinations of various machining parameters in a Computer Numerical Control (CNC) turning process. The 'If-Then' rules extracted from both the decision trees would guide the concerned process engineers in investigating the effects of the input parameters on the considered responses. Based on the detailed analysis of the corresponding decision trees and decision rules, the following conclusions can be derived:

a) For achieving higher tool life, cryogenic environment, and low values of cutting speed, tool nose radius and feed rate need to be set. Depth of cut has almost no effect on tool life;

b) A combination of cryogenic environment, low or medium cutting speed, low feed rate, and low nose radius are responsible for lower power consumption;

c) To attain lower surface roughness of the turned components, cryogenic environment, high cutting speed, low or medium feed rate, and high nose radius would be the recommended setting for the said CNC turning process. Depth of cut plays in significant role on surface roughness;

d) A parametric mix of cryogenic environment, low cutting speed, low or medium feed rate, low depth of cut, and low or medium nose radius would provide lower cutting speed;

e) The CART algorithm supersedes the CHAID algorithm with respect to higher overall classification accuracy and lower prediction risk. Although, for some of the responses, CART generates a slightly more number of decision rules as compared to CHAID, it can almost perfectly predict high as well as low values of all the responses;

f) Between these two algorithms, CART has a higher capacity to predict high values of the responses, whereas, CHAID performs better for low responses values;

g) Based on the sensitivity analysis study, the prediction performance of the CART algorithm is observed to be least affected by the perturbed response values in the experimental dataset as compared to the CHAID algorithm.

Table 9. Classification accuracies of Classification And Regression Tree (CART) and Chi-squared Automatic Interaction Detection (CHAID) algorithms at various error levels.

	Error level									
	CART					CHAID				
Response	5%	10%	15%	20%	25%	5%	10%	15%	20%	25%
TL	100	100	100	92.6	92.6	92.6	92.6	92.6	82.6	85.2
PC	100	100	96.3	96.3	92.6	100	100	92.6	96.3	96.3
SR	88.9	81.5	85.2	85.2	81.5	81.5	80.9	80.45	81.5	79.0
CF	92.6	82.6	92.6	92.6	88.9	88.9	77.8	88.9	88.9	85.2

The most preferable parametric settings for the considered CNC turning process are observed to be in close agreement with those derived by the past researchers based on Taguchi methodology, which proves the efficacy of the developed decision as revealed clearly by the study of the material removal mechanism. These classification algorithms can thus be applied to any machining process to investigate the effects of different input parameters on the responses and identify the best machining conditions for enhanced process monitoring and control.

As the decision rules mainly focus on classification, they often neglect predicting the interrelationships between the input parameters and responses in the form of regression. While a continuous variable is divided into intervals and turned into a classification problem, there is a high possibility of a loss of valuable information. Thus, there must be always a trade-off between predictive accuracy and computational effort to arrive at the most appropriate set of decision rules.

References

1. Trent, E.M., *Metal Cutting*, Woburn, Massachusetts: Butterworth-Heinemann (2010).
2. Suh, S.-H., Kang, S.K., Chung, D.-H., et al., *Theory and Design of CNC Systems*, Springer (2008).
3. Park, K.S. and Kim, S.H. "Artificial intelligent approaches to determination of CNC machining parameters in manufacturing: A review", *Artificial Intelligence in Engineering*, **12**, pp. 127–134 (1997).
4. Gupta, A., Singh, H., and Aggarwal, A. "Taguchi-fuzzy multi output optimization (MOO) in high speed CNC turning of AISI P-20 tool steel", *Expert Systems with Applications*, **38**, pp. 6822–6828 (2011).
5. Mukherjee, S., Kamal, A., and Kumar, K. "Optimization of material removal rate during turning of SAE 1020 material in CNC lathe using Taguchi technique", *Procedia Engineering*, **97**, pp. 29–35 (2014).
6. Marko, H., Simon, K., Tomaz, I., et al. "Turning parameters optimization using particle swarm optimization", *Procedia Engineering*, **69**, pp. 670–677 (2014).
7. Saini, S.K. and Pradhan, S.K. "Optimization of multi-objective response during CNC turning using Taguchi-fuzzy application", *Procedia Engineering*, **97**, pp. 141–149 (2014).
8. Vasudevan, H., Deshpande, N.C., and Rajguru, R.R. "Grey fuzzy multiobjective optimization of process parameters for CNC turning of GFRP/epoxy composites", *Procedia Engineering*, **97**, pp. 85–94 (2014).
9. Saini, S.K. and Pradhan, S.K. "Optimization of machining parameters for CNC turning of different materials", *Applied Mechanics and Materials*, **592–594**, pp. 605–609 (2014).
10. Aghdeab, S.H., Mohammed, L.A., and Ubaid, A.M. "Optimization of CNC turning for aluminum alloy using simulated annealing method", *Jordan Journal of Mechanical and Industrial Engineering*, **9**(1), pp. 39–44 (2015).
11. Camposeco-Negrete, C. "Optimization of cutting parameters using response surface method for minimizing energy consumption and maximizing cutting quality in turning of AISI 6061 T6 aluminum", *Journal of Cleaner Production*, **91**, pp. 109–117 (2015).
12. Sarıkaya, M. and Güllü, A. "Multi-response optimization of minimum quantity lubrication parameters using Taguchi-based grey relational analysis in turning of difficult-to-cut alloy Haynes 25", *Journal of Cleaner Production*, **91**, pp. 347–357 (2015).
13. Asiltürk, İ., Neşeli, S., and İnce, M.A. "Optimisation of parameters affecting surface roughness of Co28Cr6Mo medical material during CNC lathe machining by using the Taguchi and RSM methods", *Measurement*, **78**, pp. 120–128 (2016).
14. Kumar, U., Singh, A., and Kumar, R. "Optimization of machining parameters for tool wear rate and material removal rate in CNC turning by grey relational analysis", *International Journal of Applied Engineering Research*, **11**(4), pp. 2771–2775 (2016).
15. Maheswara Rao, C. and Venkatasubbaiah, K. "Optimization of surface roughness in CNC turning using Taguchi method and ANOVA", *International Journal of Advanced Science and Technology*, **93**, pp. 1–14 (2016).
16. Klancnik, S., Hrelja, M., Balic, J., et al. "Multi-objective optimization of the turning process using a gravitational search algorithm (GSA) and NSGA-II approach", *Advances in Production Engineering & Management*, **11**, pp. 366–376 (2016).
17. Bilga, P.S., Singh, S., and Kumar, R. "Optimization of energy consumption response parameters for turning operation using Taguchi method", *Journal of Cleaner Production*, **137**, pp. 1406–1417 (2016).
18. Kushwaha, A. and Singh, R. "Optimization of CNC process parameters on Inconel 625 using response surface methodology", *International Journal of Mechanical Engineering and Technology*, **8**(7), pp. 1830–1836 (2017).
19. Nataraj, M. and Balasubramanian, K. "Parametric optimization of CNC turning process for hybrid metal matrix composite", *International Journal Advanced Manufacturing Technology*, **93**, pp. 215–224 (2017).
20. Nayak, N.K. and Sodhi, H.S. "Optimization of CNC turning parameters for Al-6061 using response surface methodology", *International Journal of Mechanical and Production Engineering Research and Development*, **7**(4), pp. 127–138 (2017).
21. Sahoo, P., Pratap, A., and Bandyopadhyay, A. "Modeling and optimization of surface roughness and tool vibration in CNC turning of Aluminum alloy using hybrid RSM-WPCA methodology", *International*

Journal of Industrial Engineering Computations, **8**, pp. 385–398 (2017).

22. Mandal, N.K., Mondal, M., and Singh, N.K. "Modelling and optimisation of a sustainable manufacturing process with CNC turning centre", *International Journal of Applied Environmental Sciences*, **12**, pp. 1101–1116 (2017).

23. Suresh, M., Meby Selvaraj, R., Rajkumar, K., et al. "Optimisation of cutting parameters in CNC turning of EN-19 using tungsten carbide", *International Journal of Computer Aided Engineering and Technology*, **9**(2), pp. 218–228 (2017).

24. Akkuş, H. "Optimising the effect of cutting parameters on the average surface roughness in a turning process with the Taguchi method", *Materials and Technology*, **52**, pp. 781–785 (2018).

25. Bhanu Prakash, P., Brahma Raju, K., Venkata Subbaiah, K., et al. "Application of Taguchi based grey method for multi aspects optimization on CNC turning of $AlSi_7$ Mg", *Materials Today: Proceedings*, **5**, pp. 14292–14301 (2018).

26. Gadekula, R.K., Potta, M., Kamisetty, D., et al. "Investigation on parametric process optimization of HCHCR in CNC turning machine using Taguchi technique", *Materials Today: Proceedings*, **5**, pp. 28446–28453 (2018).

27. Palanisamy, D. and Senthil, P. "Application of grey-fuzzy approach for optimization of CNC turning process", *Materials Today: Proceedings*, **5**, pp. 6645–6654 (2018).

28. Sahoo, P., Satpathy, M.P., Singh, V.K., and Bandyopadhyay, A. "Performance evaluation in CNC turning of AA6063-T6 alloy using WASPAS approach", *World Journal of Engineering*, **15**(6), pp. 700–709 (2018).

29. Saravanakumar, A., Karthikeyan, S.C., Dhamotharan, B., et al. "Optimization of CNC turning parameters on aluminum alloy 6063 using Taguchi robust design", *Materials Today: Proceedings*, **5**, pp. 8290–8298 (2018).

30. Nataraj, M., Balasubramanian, K., and Palanisamy, D. "Optimization of machining parameters for CNC turning of Al/Al_2O_3 MMC using RSM approach", *Materials Today: Proceedings*, **5**, pp. 14265–14272 (2018).

31. Vasudevan, H., Rajguru, R., Tank, K., et al. "Optimization of multi-performance characteristics in the turning of GFRP (E) composites using principle component analysis combined with grey relational analysis", *Materials Today: Proceedings*, **5**, pp. 5955–5967 (2018).

32. Rao, V.D.P., Mahaboob Ali, S.R.S., Saqheed Ali, S.M.Z.M., et al. "Multi-objective optimization of cutting parameters in CNC turning of stainless steel 304 with TiAlN nano coated tool", *Materials Today: Proceedings*, **5**, pp. 25789–25797 (2018).

33. Vijay Kumar, M., Kiran Kumar, B.J., and Rudresha, N. "Optimization of machining parameters in CNC turning of stainless steel (EN19) by Taguchi's orthogonal array experiments", *Materials Today: Proceedings*, **5**, pp. 11395–11407 (2018).

34. Arun Vikram, K., Lakshmi, V.V.K., and Venkata Praveen, A.M. "Evaluation of process parameters using GRA while machining low machinability material in dry and wet conditions", *Materials Today: Proceedings*, **5**, pp. 25477–25485 (2018).

35. Chau, N.L., Nguyen, M.-Q., Dao, T.-P., et al. "An effective approach of adaptive neuro-fuzzy inference system-integrated teaching learning-based optimization for use in machining optimization of S45C CNC turning", *Optimization and Engineering*, **20**, pp. 811–832 (2019).

36. Susanti, Y., Zukhronah, E., Pratiwi, H., et al. "Analysis of chi-square automatic interaction detection (CHAID) and classification and regression tree (CRT) for classification of corn production", *IOP Conference Series: Journal of Physics*, **909**, pp. 1–8 (2017).

37. Rokach, L., and Maimin, O., *Data Mining with Decision Trees: Theory and Applications*, World Scientific Publishing Co., NJ, USA (2014).

38. Han, J., Kamber, M., and Pei, J., *Data Mining Concepts and Techniques*, Elsevier Inc., USA (2012).

39. Breiman, L., Friedman, J.H., Olshen, R.A., et al., *Classification and Regression Tree.*, Chapman and Hall, New York (1993).

40. Kass, G.V. "An exploratory technique for investigating large quantities of categorical data", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **29**(2), pp. 119–127 (1980).

41. Pitombo, C.S., de Souza, A.D., and Lindner, A. "Comparing decision tree algorithms to estimate intercity trip distribution", *Transportation Research Part C: Emerging Technologies*, **77**, pp. 16–32 (2017).

42. Sadoyan, H., Zakarian, A., and Mohanty, P. "Data mining algorithm for manufacturing process control", *International Journal of Advanced Manufacturing Technology*, **28**, pp. 342–350 (2006).

Biographies

Shruti Sudhakar Dandge is currently working as a Lecturer in Mechanical Engineering at Government Polytechnic, Murtizapur, India since February 2011. Presently she is pursuing a PhD in Engineering from Jadavpur University, Kolkata, India. Born in India, she received the B Eng degree (Mechanical Engineering) in 2006 from V.Y.W.S.'s Professor Ram Meghe Institute of Technology and Research, Badnera-Amravati. She qualified for GATE (Mechanical) in 2007 and holds her M-Tech degree in Thermal Engineering from Government Engineering College, Amravati in 2009. Her current research interests are the application of computational tools including statistics, machine learning, and data mining techniques in the manufacturing domain. She also carried out an industrial project named

"Thermal and Mechanical design of shell and tube heat exchangers using fixed tube-sheets" at L&T, Mumbai (India). She is interested in areas of thermodynamics, heat transfer, and analysis of I.C. Engine.

Shankar Chakraborty is a Professor in Production Engineering Department of Jadavpur University. He had been graduated in 1986 from the University of Calcutta and had obtained his postgraduate degree from Jadavpur University in 1989. He had been awarded PhD (Engg.) degree from Jadavpur University in 1994. His research interests include applications of different multi-criteria decision-making methods in a manufacturing environment, control chart pattern recognition, and development of MIS and ERP systems for diverse engineering applications, soft computing techniques etc. He has guided several ME and PhD (Engg.) theses and published numerous papers in international journals. He is also a regular reviewer of several journals of international repute.

Management and Production Engineering Review
Volume 12 • Number 3 • September 2021 • pp. 116–128
DOI: 10.24425/mper.2021.138536

A Data Mining Approach for Analysis of a Wire Electrical Discharge Machining Process

Shruti Sudhakar Dandge[1], Shankar Chakraborty[2]

[1] *Mechanical Engineering Department, Government Polytechnic, Murtizapur, Maharashtra, India*
[2] *Department of Production Engineering, Jadavpur University, Kolkata, West Bengal, India*

Received: 18 November 2019
Accepted: 27 May 2021

Abstract
Wire electrical discharge machining (WEDM) is a non-conventional material-removal process where a continuously travelling electrically conductive wire is used as an electrode to erode material from a workpiece. To explore its fullest machining potential, there is always a requirement to examine the effects of its varied input parameters on the responses and resolve the best parametric setting. This paper proposes parametric analysis of a WEDM process by applying non-parametric decision tree algorithm, based on a past experimental dataset. Two decision tree-based classification methods, i.e. classification and regression tree (CART) and Chi-squared automatic interaction detection (CHAID) are considered here as the data mining tools to examine the influences of six WEDM process parameters on four responses, and identify the most preferred parametric mix to help in achieving the desired response values. The developed decision trees recognize pulse-on time as the most indicative WEDM process parameter impacting almost all the responses. Furthermore, a comparative analysis on the classification performance of CART and CHAID algorithms demonstrates the superiority of CART with higher overall classification accuracy and lower prediction risk.

Keywords
wire electrical discharge machining, data mining, classification and regression tree, chi-squared automatic interaction detection, classification.

Introduction

In present day manufacturing industries, wire electrical discharge machining (WEDM) process has become quite popular due to its competence to machine various tough and hard-to-machine materials with complex shape geometries and close tolerances. The WEDM is a variant of electrical discharge machining (EDM) process where a continuously travelling electrically conductive wire (e.g. tungsten, brass or copper with diameter between 0.05 and 0.3 mm) is used as the electrode. This wire is kept in tension using a mechanical device and its movement is numerically controlled so as to attain the desired accuracy and tolerance while machining a given workpiece. In WEDM process, removal of material usually occurs due to complex erosion effect of expeditious, continual and distinct spark discharges between the wire tool and the workpiece submerged in kerosene or deionized water (dielectric medium). These electrical discharges cause melting and vaporization of tiny amounts of material from the workpiece, which are rinsed away by the dielectric, causing small pits on the workpiece. In WEDM process, as the material is abraded before the wire and as there is no direct contact between the workpiece and the wire, chances of generation of stress, chatter and vibration during the machining operation are less (Ho et al., 2004; Mandal and Dixit, 2014). As WEDM is distinguished to be a highly accurate process, it has now found wide-ranging applications in aerospace, nuclear, automotive, bio-medical, and tool and die-making industries. It can machine various unusual high-strength-temperature-resistive materials, like alloys, titanium, cemented carbides, ceramics and silicon (Patel and Vaghmare, 2013). In a manufacturing industry, the main goal of WEDM application is to realize higher machining rate (MR) along with better dimensional accurateness and surface characteristic. However, the performance of a WEDM process re-

Corresponding author: *Shankar Chakraborty – Department of Production Engineering, Jadavpur University, Kolkata, West Bengal, India, phone: +91 983 156 8294, e-mail: s_chakraborty00@yahoo.co.in*

garding material removal rate (MRR), surface roughness (SR), wire wear ratio (WWR), kerf width, dimensional deviation (DD) etc. is often observed to be influenced by several controllable (input) parameters, like peak current (I_p) (in A), pulse-on time (T_{on}) (in μs), pulse-off time T_{off}) (in μs), wire tension (WT) (in g), wire feed rate (WF) (in m/min), spark gap voltage (SV) (in V) etc. The presence of an extensive set of process parameters, stochastic nature of the process, possible interactions between the process parameters and conflicting behaviour of the responses make it imperative to explore the effects of various WEDM process parameters on the outputs. Thus, to accomplish the desired response values, identification of the optimal machining parameters and their settings plays a significant role. It has been observed that inaccurately selected WEDM process parameters may often lead to short-circuiting of wire, breakage of wire and damage of the workpiece surface. With the rapidly growing use of WEDM process for machining newer and advanced engineering materials, an ardent need is thus acknowledged for development and deployment of an innovative approach for studying the impacts of varied process parameters on the responses and identifying the parametric mix leading to optimization of WEDM process. From the review of the contemporary literature (Kumar et al., 2013; Lal et al., 2015; Dabade and Karidkar et al., 2016; Lusi et al., 2016; Manjaiah et al., 2016; Arikatla et al., 2017; Ramanan and Elangovan, 2018; Devarajaiah and Muthumari, 2018; Vignesh and Ramanujam, 2018; Srinivasarao and Suneel, 2018; Nayak et al., 2018), it has been revealed that determination of the optimal parametric mixes for WEDM processes has already caught the attention of the past researchers, and numerous multi-objective optimization techniques, like grey-Taguchi method, desirability function approach, technique for order of preference by similarity to ideal solution (TOPSIS), evolutionary algorithms etc. have been applied. But, the applications of various data mining tools for studying the influences of varying WEDM process parameters on the responses and identifying the optimal combinations of those process parameters are really limited. Thus, this paper focuses on the application of data mining techniques, primarily through decision trees, to fulfil the above-mentioned objectives. Two decision tree-based classification algorithms, i.e. classification and regression tree (CART) and Chi-squared automatic interaction detection (CHAID) are thus employed here, and their performance is contrasted with respect to overall classification accuracy and prediction risk.

Data mining

Data mining, occasionally known as 'knowledge discovery in databases', is the procedure of drilling through data to unveil hidden patterns or connections and anticipate future trends (Tan et al., 2006; Han et al., 2012). It mainly depends on effective collection of data and warehousing as well as computer processing. The application of data mining has strong relations with statistics (to study data relationships), artificial intelligence (to provide human-like intelligence) and machine learning (to learn from data to make predictions). Data mining thus allows to (a) filter all the chaotic and repetitive noise in the dataset, (b) predict automatic pattern based on trend and behaviour analysis, (c) envisage the likely outcomes, (d) create decision-oriented information, (e) accelerate the pace of making informed decisions, (f) focus on large dataset and database for analysis, and (g) form clusters based on visually documented groups of facts. In predictive modelling, the primary objective is to approximate the value of a specific target attribute from a set of training data where the attribute values are established beforehand. Classification is a unique example of predictive modelling where a data set is already segmented into pre-specified groups and patterns are identified in the data to distinguish those groups. The explored patterns can then be adopted to categorize another dataset where the appropriate group description for the target attribute is unknown. Regression analysis is also an example of predictive modelling with numerical target attribute and the objective is to envision that value for new data. The application of data mining techniques follows some procedural steps, such as (a) data cleaning (elimination of noisy and outlier data), (b) data assimilation (amalgamation of multiple data sources), (c) data selection (recovery of pertinent data from the database), (d) data conversion (conversion of data for mining purposes), (e) data mining (identification of data patterns), (f) pattern assessment (extraction of attractive patterns) and (g) knowledge display (picturing of the mined information) (Ramani et al., 2020).

Decision tree analysis

The decision tree analysis can be defined as a set of practices to estimate and show the extant relations between a dependent variable and a class of independent variables. It is based on consecutive partitioning algorithm to continuously split the data to constitute homogeneous subsets, producing a hierarchical tree consisting of decision rules convenient for data anticipation or classification. The main classification methods for decision tree induction are CART and CHAID algorithms, and their characteristics are outlined as below:

1. The resultant hierarchy is known as a tree and a particular section is referred to as a node.
2. The root node consists of the entire database.
3. It is branched conclusively forming child nodes.
4. The final subgroups are known as terminal nodes or leaves when no further data classification is possible.
5. There are three primary elements to be defined to implement them, i.e. a set of questions delimitating data apportionment, a criterion to institute the best division to develop child nodes and a completion rule for the classifications (stop-splitting rule).

Both of them build decision trees, where each (non-terminal) node establishes a split condition to produce an optimal prediction for continuous dependent variables or classification for categorical dependent variables.

CART-based algorithm

Breiman et al. (1993) proposed this classification algorithm for origination of decision trees. It divides a particular population into binary splits, while consecutively breaking up the data into smallest components with maximal homogeneity with regard to the dependent variable. It follows a sequential procedure, as enlisted below:

1. Tree growing process – This algorithm evaluates all the possible splits of the explanatory variables and designates the 'best' split, beginning with the root node. Thereafter, this process is imitated for consequent nodes. The 'best' split is explained by the minimal overall impurity. Usually, in this algorithm, univariate splits are only designed. Hence, beginning with the root node, a tree is developed in the downward direction by repeatedly performing the splitting operation.
2. Splitting criterion and impurity measure – At a particular node, the 'best' split is considered for maximization of a specific splitting criterion. For a specific impurity measure for a node, the splitting criterion conforms to decrement in impurity. The 'Gini' index is adopted as the impurity function in this algorithm. The chosen independent variable is that one which assures a segmentation having the maximum level of improvement.
3. Stopping rule – The tree growing process would stop while satisfying any of the following stopping rules:
   i) When all cases in a particular node have same values of the dependent variable.
   ii) When the present tree depth arrives the maximum limit as stated by the user.
   iii) When the node size is smaller than the minimum size as stated by the user.
   iv) When the node splitting forms a child node having size smaller than the user-identified minimum size.
4. Variable importance – The importance measure of an explanatory variable X in the developed tree is stated as the sum across all the splits in the tree showing the gains that X achieves when it is adopted as a dominant or surrogate splitter. The importance of variable X is denoted with respect to a normalized quantity as compared to the variable having the maximum measure of importance. Its value spans between 0 and 100, having the variable with the largest measure of importance score as 100. Thus, an explanatory variable's importance is a preferred indicator to account the significance of disaggregated variable as well as aggregated variable (which has already appeared in the decision tree).

CHAID-based algorithm

This algorithm, proposed by Kass (1980), generates non-binary decision trees (with more than two branches connected to a single node) based on the Bonferroni's test. It proceeds through the following three stages:

1. Merging – For each dependent variable X, combine the non-significant categories. Each final category of X results in one child node if X is utilized to divide the node. This merging step also estimates the revised p-value to be subsequently used for splitting purpose.
2. Splitting – The explanatory variable having the lowest significant p-value is identified as the best and the group is divided based on this predictor. The group is not divided when no predictor has a significant p-value.
3. Stopping – These stages are imitated till all the subgroups have either been inspected or have encompassed too few observations. The adopted stopping rules are mostly the same as those already mentioned for CART algorithm.

Development of decision trees for WEDM process

In a four-axis CNC WEDM set-up, Kumar et al. (2013) performed 54 experiments to investigate the effects of six process parameters, e.g. T_{on}, T_{off}, I_P, SV, WF and WT on four responses. The responses were MR (in mm/min), SR (in µm), DD (in µm) and WWR. Other factors, like type of the workpiece material (pure titanium-grade 2), wire electrode (brass wire having diameter 0.25 mm), workpiece thickness and pressure of the dielectric were kept constant during the experimental runs. Table 1 exhibits the detailed experimental plan and measured values of the responses. Among these responses, MR is the sole beneficial quality characteristic desired with its higher value. On the contrary, minimum values are required for SR, DD and WWR (non-beneficial attributes (Sarker and Chakraborty, 2021).

Table 1
Experimental observations for the WEDM process (Kumar et al., 2013)

Run No.	I_p	T_{on}	SV	T_{off}	WF	WT	SR	WWR	MR	DD
1	200	120	50	50	7	500	3.22	0.095	1.14	160
2	160	116	50	56	4	500	2.48	0.063	0.576	150
3	160	112	60	50	4	950	2.23	0.079	0.42	145
4	120	116	50	44	10	950	2.75	0.086	0.954	159
5	120	116	60	50	7	500	2.47	0.061	0.544	152
6	160	120	40	50	4	950	2.93	0.088	1.075	162
7	160	116	50	56	10	1400	2.48	0.063	0.586	150
8	160	116	50	50	7	950	2.65	0.080	0.695	152
9	160	116	50	44	4	500	2.81	0.089	1.014	160
10	160	120	40	50	10	950	2.94	0.088	1.075	160
11	160	120	40	56	7	950	2.91	0.087	0.995	160
12	160	120	60	50	4	950	2.83	0.079	0.809	159
13	160	116	50	44	10	500	2.79	0.076	1.012	160
14	160	116	50	50	7	950	2.61	0.064	0.573	150
15	120	112	50	50	7	500	2.49	0.048	0.406	145
16	160	116	50	50	7	950	2.68	0.082	0.697	152
17	120	116	60	50	7	1400	2.49	0.059	0.538	150
18	160	112	40	56	7	950	2.32	0.060	0.48	145
19	120	116	50	56	10	950	2.31	0.056	0.535	151
20	200	116	40	50	7	1400	2.89	0.079	0.825	152
21	200	116	60	50	7	500	2.69	0.072	0.773	152
22	200	116	50	56	10	950	2.57	0.074	0.792	153
23	120	116	40	50	7	1400	2.71	0.068	0.625	152
24	120	112	50	50	7	1400	2.51	0.054	0.425	145
25	200	116	50	56	4	950	2.56	0.078	0.799	155
26	160	120	60	50	10	950	2.82	0.081	0.81	153
27	120	120	50	50	7	500	2.77	0.074	0.83	158
28	160	112	40	50	10	950	2.35	0.085	0.521	150
29	200	112	50	50	7	500	2.48	0.083	0.535	150
30	160	112	40	44	7	950	2.70	0.089	0.858	153
31	200	112	50	50	7	1400	2.51	0.082	0.54	150
32	160	116	50	50	7	950	2.65	0.081	0.658	150
33	200	116	50	44	4	950	2.88	0.092	1.02	159
34	160	116	50	50	7	950	2.65	0.081	0.656	152
35	160	120	40	44	7	950	3.28	0.107	1.28	165
36	200	116	50	44	10	950	2.98	0.095	1.03	160
37	200	116	40	50	7	500	2.84	0.079	0.829	155
38	160	112	40	50	4	950	2.33	0.081	0.529	150
39	160	116	50	56	10	500	2.50	0.064	0.589	150
40	160	116	50	50	7	950	2.69	0.081	0.659	152
41	160	120	60	56	7	950	2.66	0.070	0.792	153
42	160	112	60	44	7	950	2.60	0.081	0.495	150
43	200	116	60	50	7	1400	2.68	0.072	0.778	155
44	120	116	50	44	4	950	2.75	0.086	0.959	155
45	160	112	60	50	10	950	2.28	0.079	0.429	145
46	120	120	50	50	7	1400	2.75	0.074	0.823	158
47	160	112	60	56	7	950	2.15	0.064	0.395	140
48	160	116	50	44	4	1400	2.85	0.088	0.981	159
49	120	116	40	50	7	500	2.78	0.068	0.635	158
50	160	120	60	44	7	950	3.00	0.085	1.00	159
51	120	116	50	56	4	950	2.29	0.060	0.541	150
52	200	120	50	50	7	1400	3.12	0.091	1.052	159
53	160	116	50	44	10	1400	2.82	0.088	0.962	155
54	160	116	50	56	4	1400	2.49	0.060	0.592	150
Median							0.74	2.68	152	0.079

In this paper, CART and CHAID algorithms are applied to explore the experimental dataset of Kumar et al. (2013) while generating the corresponding decision trees and induction rules to study the influences of the considered WEDM process parameters on the four responses. An endeavour is also put forward to compare their relative classification performance with respect to prediction accuracy and prognosis risk. In a decision tree, an internal node denotes a test on an attribute, a branch characterizes the result of the test and a leaf node depicts a class label. A specific decision or induction rule can thus be obtained while following the direction from the root to leaf node. For development of the related decision trees using CART and CHAID algorithms (available in SPSS 16.0), the following specifications are predefined.

For CART algorithm:

Growing method: CART, Categorical dependent variables: MR, SR, DD and WWR, Continuous independent variables: T_{on}, T_{off}, I_P, SV, WF and WT, Number of sample folds: 3, Validation: Cross validation, Growth limit: Maximum depth of tree = 5, Minimum number of cases: Parent node = 3, Child node = 2, Impurity measure: 'Gini', Minimum shift in improvement = 0.0001.

For CHAID algorithm:

Growing method: CHAID, Categorical dependent variables: MR, SR, DD and WWR, Categorical independent variables: T_{on}, T_{off}, I_P, SV, WF and WT,

Validation: Cross validation, Growth limit: Maximum depth of tree = 5, Number of sample folds: 3, Minimum number of cases: Parent node = 3, Child node = 2, Significance level for: a) Splitting node = 0.03, b) Merging categories = 0.05 and c) Chi-square statistic = Pearson, Model estimation: a) Maximum number of iterations = 100, b) minimum variation in expected cell frequencies = 0.001 and c) Modify significance values based on the Bonferroni method.

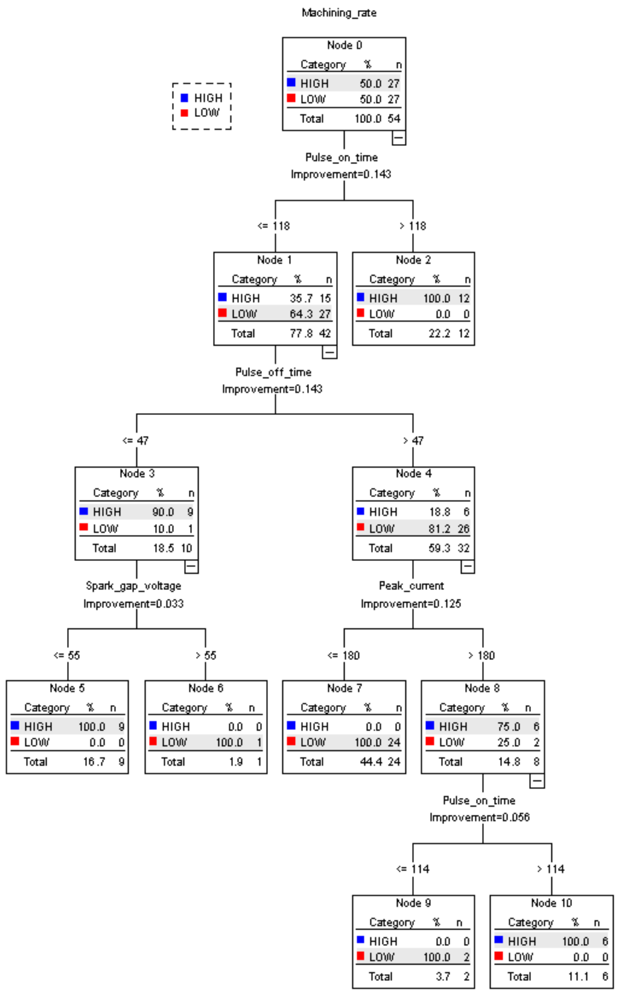


Fig. 1. Decision tree depicting the impacts of WEDM process parameters on MR

Figure 1 exhibits the CART algorithm-based decision tree, showing the impacts and contributions of the six WEDM process parameters on MR. In this figure, the term 'low' belongs to observations with MR $\leq$ 0.74 mm/min and 'high' refers to those observations having MR $>$ 0.74 mm/min (where 0.74 mm/min is the median value of MR). This classification tree contains 5 splits and 6 terminal nodes. The classification process begins with the top or root node with level 0. All the 54 observations are first assigned to this node where 27 observations are equally classified as 'low' and 'high', as represented in this node. The root node is again divided into two new nodes and the related split condition is shown below the root node. It can be noticed from this decision tree that during the first classification, 42 experimental observations with T_{on} less than or equal to 118 µs are routed to node number 1, categorized as 'low' items. The remaining 12 observations with T_{on} greater than 118 µs are directed to node number 2 with 'high' classification. All the 12 observations belonging to only 'high' value identify node 2 as a pure node with no misclassification error. In node 1, unequal number of observations recognizes it as an impure node providing more child nodes. Thus, node 1 is now split giving rise to two impure nodes. In this split, 10 observations with T_{off} less than or equal to 47 µs are dispatched to node number 3, and the rest 32 observations with T_{off} greater than 47 µs are directed to node number 4. Now, based on the values of SV, node 3 is consequently divided into two pure nodes. In this split, nine observations with SV less than or equal to 55 V are attached to node number 5 with 'high' values of classification, and only one observation with SV greater than 55 V is routed to node number 6 with 'low' classification value. Similarly, another splitting is carried out from node 4 based on I_P. In this split, 24 observations with I_P less than or equal to 180 A are dispatched to node number 7 with 'low' categorization, and the remaining eight observations with I_P greater than 180 A are sent to node number 8. From node number 8, two pure nodes again emerge out based on the values of

T_{on}. In this splitting operation, two observations having T_{on} less than or equal to 114 µs are routed to node number 9 and the rest six observations with T_{on} greater than 114 µs are sent to node number 10. In this decision tree, it is noticed that all the terminal nodes are pure representing no misclassification error. The percentage of correct classification at each of the terminal nodes is presented in Table 2. Now, from this tree, the following decision/induction rules can be structured to envisage the impacts of various WEDM process parameters on MR.

Table 2
Percentage of correct classification of MR based on CART algorithm

Classification	Terminal node	MR			
		Low ($\leq$ 0.74 mm/min)		High (> 0.74 mm/min)	
		Number of observations	%	Number of observations	%
1	Node 2	0	0	12	100
2	Node 5	0	0	9	100
3	Node 6	1	100	0	0
4	Node 7	24	100	0	0
5	Node 9	2	100	0	0
6	Node 10	0	0	6	100

Decision rules for MR based on CART:
Rule 1: If $T_{on} > 118$ µs Then MR is (0.74–1.28]
$[P = 100\%,\ Q = 44.44\%,\ C = 22.22\%,\ QTY = 12]$
$[T = 166.67\%]$
Rule 2: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV ≤ 55 V Then MR is [0.74–1.28]
$[P = 100\%,\ Q = 33.33\%,\ C = 16.67\%,\ QTY = 9]$
$[T = 150\%]$
Rule 3: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV > 55 V Then MR is [0.40–0.74]
$[P = 100\%,\ Q = 3.70\%,\ C = 1.85\%,\ QTY = 1]$
$[T = 105.55\%]$
Rule 4: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P \leq 180$ A Then MR is [0.40–0.74]
$[P = 100\%,\ Q = 88.88\%,\ C = 44.44\%,\ QTY = 24]$
$[T = 233.32\%]$
Rule 5: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P > 180$ A and $T_{on} \leq 114$ µs Then MR is [0.40–0.74]
$[P = 100\%,\ Q = 7.41\%,\ C = 3.70\%,\ QTY = 2]$
$[T = 111.11\%]$
Rule 6: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P > 180$ A and $T_{on} > 114$ µs Then MR is [0.74–1.28]
$[P = 100\%,\ Q = 22.22\%,\ C = 11.11\%,\ QTY = 6]$
$[T = 133.33\%]$,

where P is the rule confidence, Q is the percentage of items in current equivalence class conforming to a rule, C is the rule support, QTY is the number of observations corresponding to a rule and T (total strength) = (P + Q + C) (Agarwal et al., 2019).

Among the developed decision rules, rule 4 with the maximum total strength of 233.32 states that in the considered WEDM process, when T_{on} is less than or equal to 118 µs, T_{off} is greater than 47 µs and I_P is less than or equal to 180 A, the corresponding MR would be low. On the other hand, rule 1 with a total strength 166.67 reveals that high T_{on} leads to high MR. As MR is a beneficial response, it is thus advised to operate the WEDM process at high setting of T_{on} (above 118 µs).

In CART-based decision tree, where univariate splits are considered, the predictor variables are usually rated on a 0–100 scale depending on their preference in accounting for the responses on the dependent variable (MR). It can be noticed from Fig. 2 that T_{on} has the maximum influence on MR response, followed by T_{off}, I_P and SV. In this WEDM process, MR is observed to be totally unaffected by WF and WT.

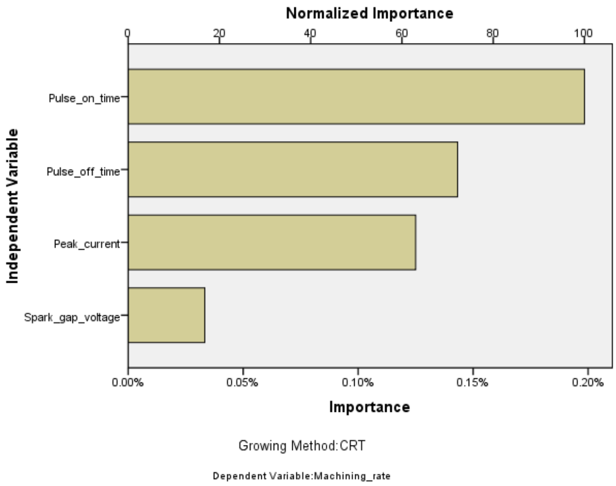


Fig. 2. Importance of various WEDM process parameters on MR

Similarly, the developed decision tree for MR using CHAID algorithm is exhibited in Fig. 3. As compared to CART algorithm, same number of classifications is also obtained in CHAID algorithm. In this multi-split decision tree, node 7 contains non-homogenous observations, identifying it as an impure node with 1.85% misclassification error. The corresponding decision rules identify T_{on}, T_{off} and I_P as the most indicative WEDM process parameters influencing MR, whereas, SV, WF and WT appear to be insignificant parameters.

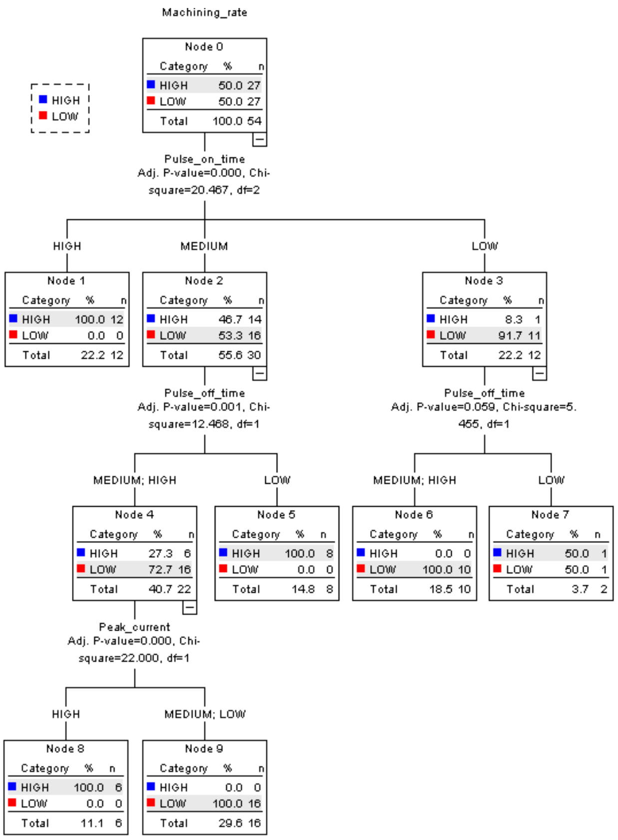


Fig. 3. CHAID algorithm-based decision tree for MR

Decision rules for MR based on CHAID:
Rule 1: If T_{on} = High Then MR is (0.74-1.28]
[P = 100%, Q = 44.44%, C = 22.22%, QTY = 12]
[T = 166.66%]
Rule 2: If T_{on} = Medium and T_{off} = Low Then MR is [0.74–1.28]
[P = 100%, Q = 29.62%, C = 14.81%, QTY = 8]
[T = 144.43%]
Rule 3: If T_{on} = Low and T_{off} = Low Then MR is [0.74–1.28]
[P = 100%, Q = 37.04%, C = 18.52%, QTY = 10]
[T = 155.56%]
Rule 4: If T_{on} = Low and T_{off} = Medium or High Then MR is [0.40–0.74]
[P = 50%, Q = 3.70%, C = 1.85%, QTY = 1]
[T = 55.55%]
Rule 5: If T_{on} = Medium and T_{off} = Medium or High and I_P = High Then MR is [0.74–1.28]
[P = 100%, Q = 22.22%, C = 11.11%, QTY = 6]
[T = 133.33%]
Rule 6: If T_{on} = Medium and T_{off} = Medium or High and I_P = Medium or Low Then MR is [0.40–0.74]
[P = 100%, Q = 59.26%, C = 29.63%, QTY = 16]
[T = 188.89%]

The CART algorithm-based decision tree representing the influences of varied WEDM process parameters on SR is depicted in Figure 4. In this case, the measured SR values are also assorted into two classes, i.e. 'low' containing SR values ≤ 2.68 µm and 'high' has SR values > 2.68 µm (where 2.68 µm is the median SR value). The induction rules formulated from the developed decision tree state that when T_{on} is greater than 118 µs, the resulting SR would be high. It can also be highlighted from the developed rules that when T_{on} is less than or equal to 118 µs, T_{off} is greater than 47 µs and SV is greater than 45 V, the SR would be low. The rules extracted from the decision tree developed using CHAID algorithm, as exhibited in Fig. 5, also prove that high T_{on} results in high SR. Medium or low T_{off} is also responsible for

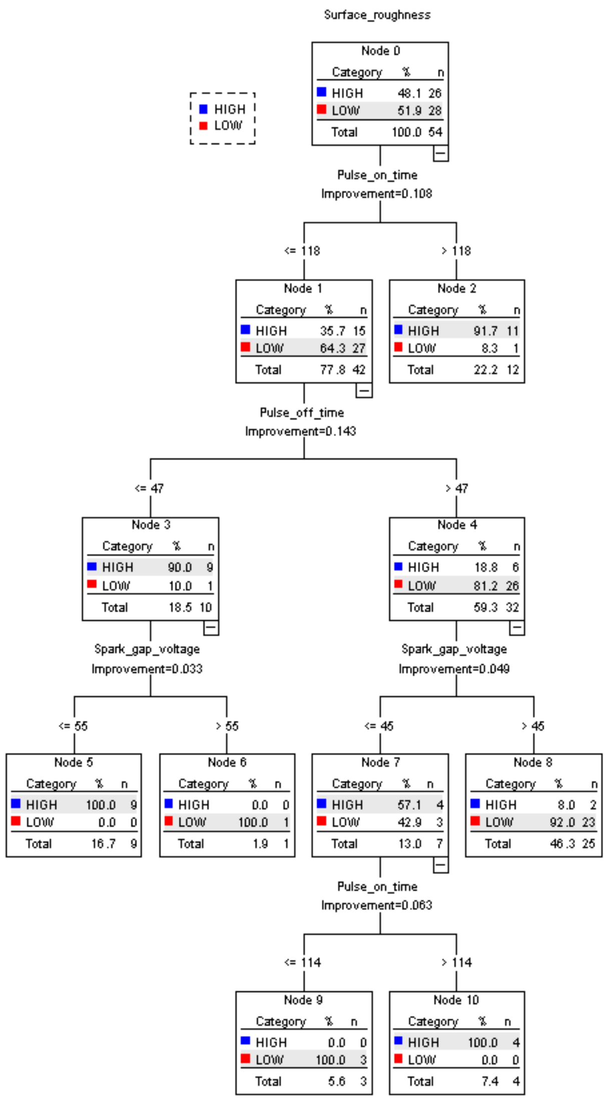


Fig. 4. Decision tree for SR using CART algorithm

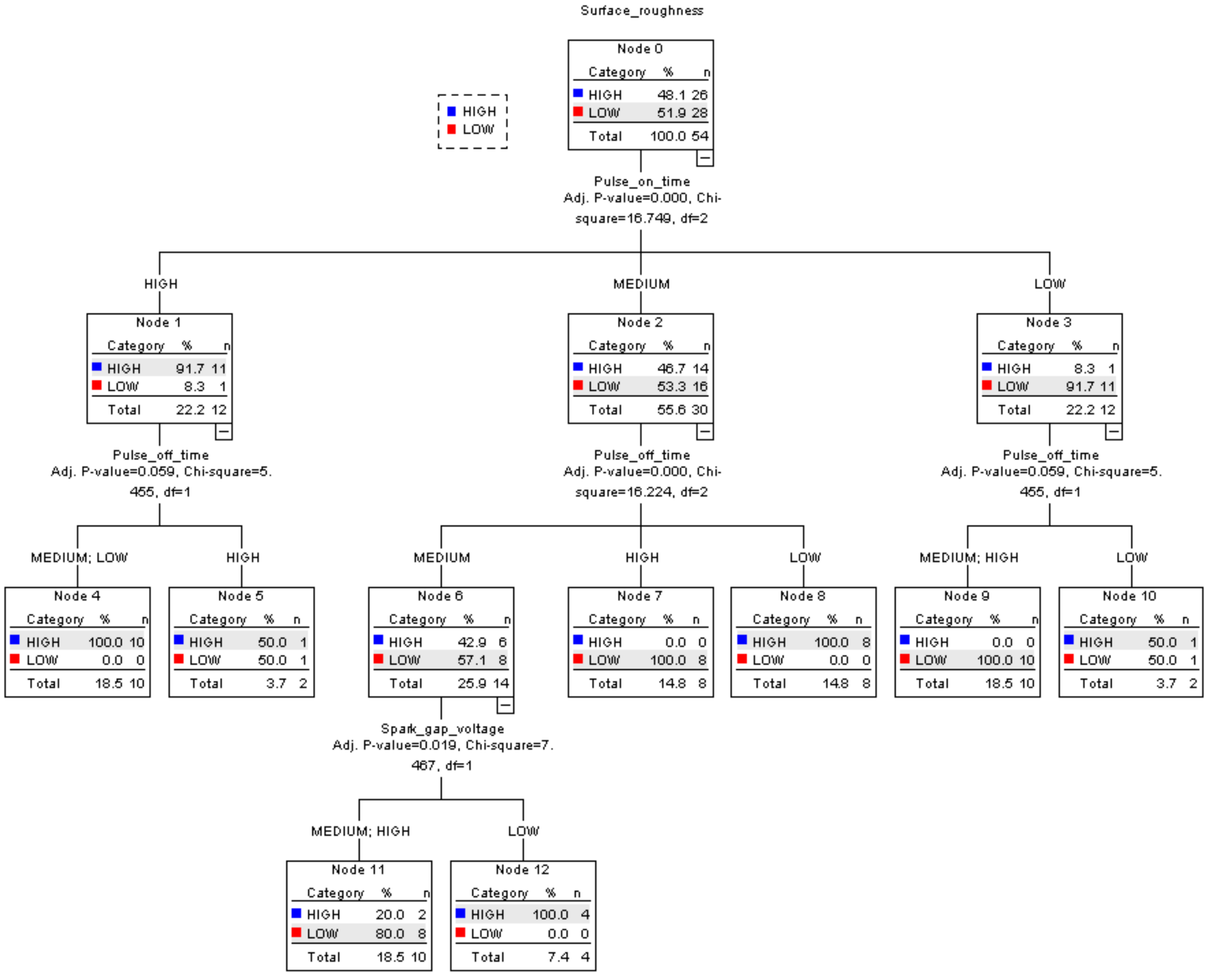


Fig. 5. CHAID algorithm-based decision tree for SR

increased SR. Similarly, low SV is responsible for high value of SR. In both the sets of rules developed using CART and CHAID algorithms, WF and WT appear to be irrelevant WEDM process parameters having no influences on SR.

Figure 6 depicts the relative importance of various WEDM process parameters on SR. It can be revealed from this figure that T_{on} plays the dominant role in controlling SR of the machined components, followed by T_{off} and SV. The other three process parameters, i.e. WF, I_P and WT have less importance on SR.

To visualize the effects of various WEDM process parameters on DD response, the decision tree is now generated using CART algorithm. When the values of DD are less than or equal to 152 µm, they are denoted as 'low' and when its values are greater than 152 µm, they are designated as 'high' (where 152 µm is the median of DD). The 'If-Then' rules extracted from this decision tree highlight that T_{on} less than or equal to 118 µs, T_{off} greater than 47 µs and I_P less than or equal to 180 A always lead to lower DD. Higher T_{on} is liable for higher DD. Low T_{on}, high T_{off} and low SV would cause lower DD.

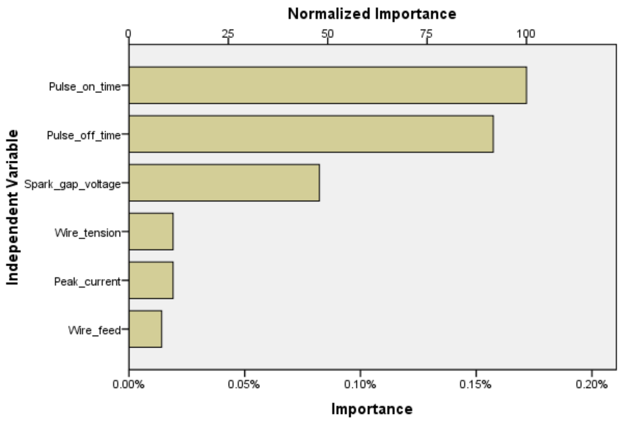


Fig. 6. Importance of various WEDM process parameters on SR

The induction rules generated based on CHAID algorithm also confirm these observations. Combining both the sets of decision rules from CART and CHAID algorithms, the optimal parametric mix of low T_{on}, medium or high T_{off}, low I_P and high SV would always lead to lower DD. Kumar et al. (2013) also observed that almost the same combination of the input parameters would be responsible for attaining lower DD in the said WEDM process. The importance plot of Fig. 7 identifies T_{on} as the most critical WEDM process parameter affecting DD, followed by T_{off}, I_P and SV. Interestingly, WF and WT have no significant roles in controlling the DD of the machined components.

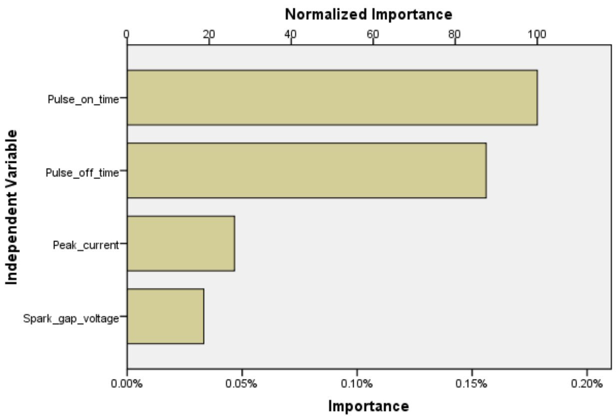


Fig. 7. Importance of varied WEDM process parameters on DD

Decision rules for DD using CART:
Rule 1: If $T_{on} > 118$ µs Then DD is [152–165]
$[P = 100\%,\ Q = 46.15\%,\ C = 22.22\%,\ QTY = 12]$
$[T = 168.37\%]$
Rule 2: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV ≤ 55 V Then DD is [152–165]
$[P = 100\%,\ Q = 34.61\%,\ C = 16.67\%,\ QTY = 9]$
$[T = 151.28\%]$
Rule 3: If $T_{on} \leq 118$ µs and $T_{off} \leq 47$ µs and SV > 55 V Then DD is [140–152]
$[P = 100\%,\ Q = 3.57\%,\ C = 1.85\%,\ QTY = 1]$
$[T = 105.42\%]$
Rule 4: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P \leq 180$ A Then DD is [140–152]
$[P = 95.80\%,\ Q = 82.14\%,\ C = 52.60\%,\ QTY = 23]$
$[T = 230.54\%]$
Rule 5: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P > 180$ A and $T_{on} > 114$ µs Then DD is [152–165]
$[P = 66.67\%,\ Q = 15.38\%,\ C = 7.41\%,\ QTY = 4]$
$[T = 89.46\%]$
Rule 6: If $T_{on} \leq 118$ µs and $T_{off} > 47$ µs and $I_P > 180$ A and $T_{on} \leq 114$ µs Then DD is [140–152]
$[P = 100\%,\ Q = 7.14\%,\ C = 3.70\%,\ QTY = 2]$
$[T = 110.84\%]$

The decision tree for WWR originated using CART algorithm is shown in Fig. 8. The corresponding

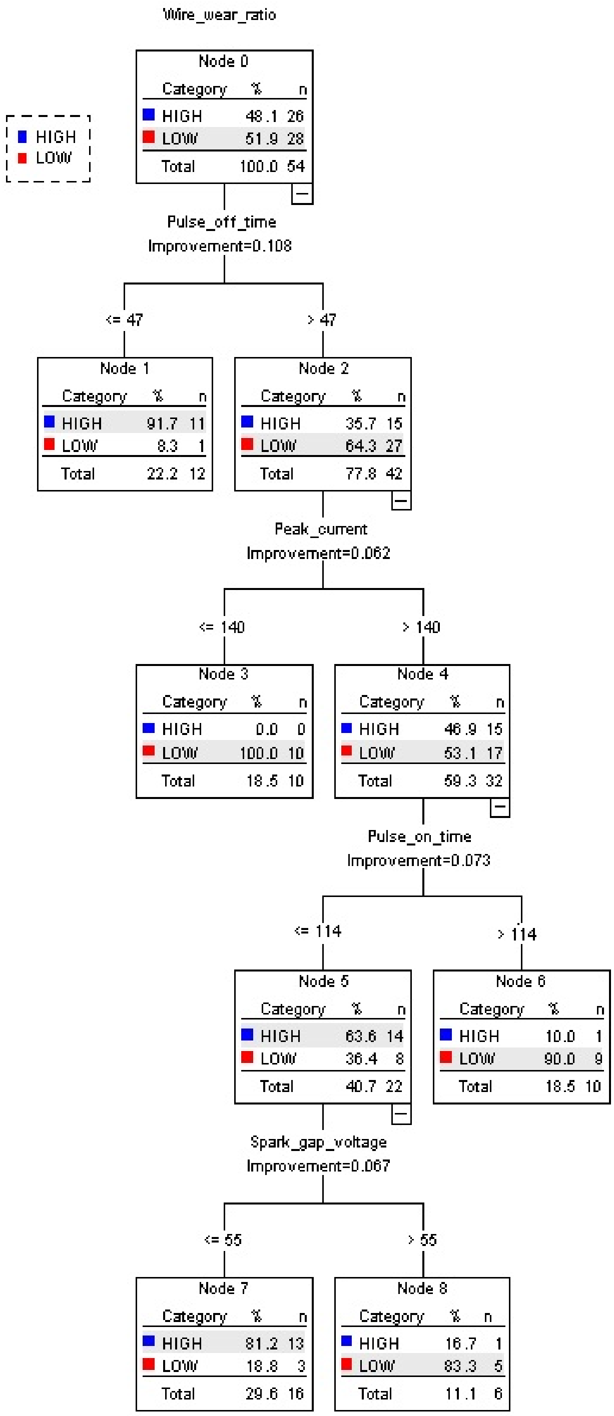


Fig. 8. Decision tree for WWR using CART algorithm

'If-Then' rules are subsequently generated from it. When the values of WWR are less than or equal to 0.079, they are denoted as 'low' and when its values are greater than 0.079, they are designated as 'high' (where 0.079 is the median of WWR. An analysis of these rules reveals that when T_{off} time is greater than 47 µs and I_P is less than or equal to 140 A, the achievable WWR is low. Similarly, lower T_{off} leads to higher WWR. The rules extracted from the decision tree based on CHAID algorithm, as shown in Fig. 9, state that T_{off} would considerably affect WWR. On the contrary, WT is least responsible for attaining lower WWR. From Fig. 10, T_{off} is identified as the most indicative parameter influencing WWR, followed by SV, I_P and T_{on}. WF and WT appear to be insignificant WEDM process parameters for WWR.

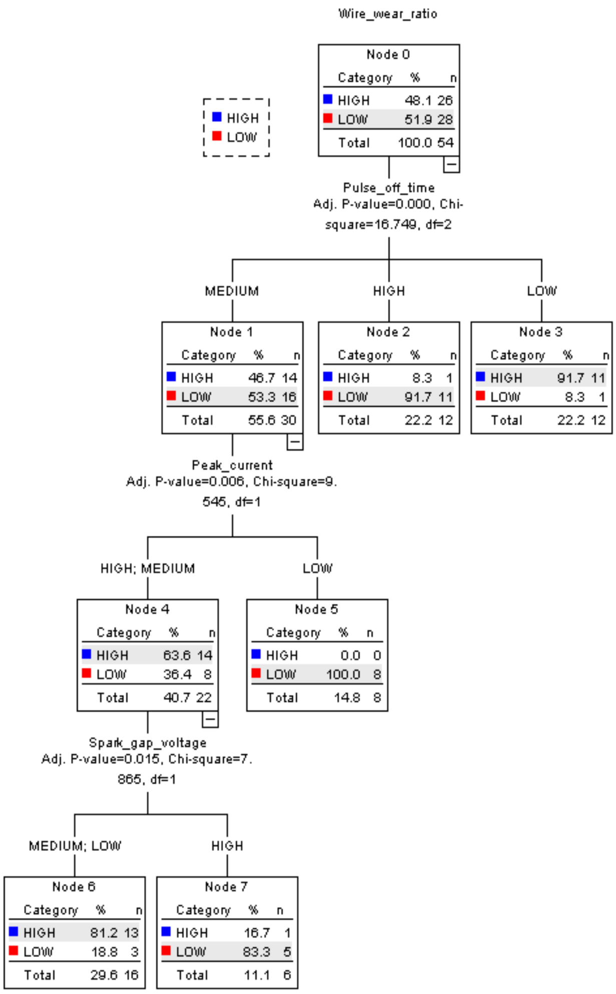


Fig. 9. CHAID algorithm-based decision tree for WWR

An optimal parametric mix of low or moderate T_{off}, high I_P, low or moderate T_{on} and moderate SV was identified by Kumar et al. (2013) for attaining lower WWR, which almost corroborates the decision tree-based observations.

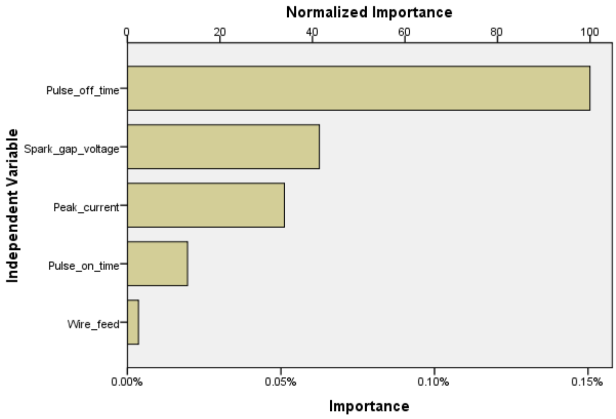


Fig. 10. Importance of WEDM process parameters on WWR

In Table 3, a comparison of the classification accuracies of CART and CHAID algorithms for all the four WEDM responses is provided. From this table, it can be noted that for MR response, CART algorithm can perfectly predict both of its low and high values. For this algorithm, the classification accuracies for high and low SR values are 92.3% and 96.4% respectively. Similarly, using CART algorithm, high and low DD values can be predicted with accuracies of 96.2% and 92.9% respectively. For WWR response, CART algorithm can respectively predict the corresponding high and low values with 92.30% and 89.28% accuracies. Thus, CART algorithm can almost perfectly predict both the high and low values of all the WEDM responses, although it has a slightly greater likelihood to accurately envisage high values of the responses.

In case of CHAID algorithm, high values of MR and SR are respectively predicted with 100% and 92.3% accuracies. It has prediction accuracies of 96.2% and 89.3% respectively for high and low DD values. For this algorithm, the classification accuracy for low WWR is the least (85.7%). Based on the overall classification accuracies of both these algorithms, it can be concluded that CART algorithm outperforms CHAID algorithm for all the four responses.

Table 3
Classification accuracies for MR, SR, DD and WWR using CART and CHAID algorithms

Response	CART				CHAID			
MR		Predicted				Predicted		
	Observed	High (> 0.74 mm/min)	Low (≤ 0.74 mm/min)	Percent correct	Observed	High (> 0.74 mm/min)	Low (≤ 0.74 mm/min)	Percent correct
	High (> 0.74 mm/min)	27	0	100	High (> 0.74 mm/min)	27	0	100
	Low (≤ 0.74 mm/min)	0	27	100	Low (≤ 0.74 mm/min)	1	26	96.3
	Overall percentage	50	50	100	Overall percentage	51.9	48.1	98.1
SR		Predicted				Predicted		
	Observed	High (> 2.68 µm)	Low (≤ 2.68 µm)	Percent correct	Observed	High (> 2.68 µm)	Low (≤ 2.68 µm)	Percent correct
	High (> 2.68 µm)	24	2	92.3	High (> 2.68 µm)	24	2	92.3
	Low (≤ 2.68 µm)	1	27	96.4	Low (≤ 2.68 µm)	2	26	92.9
	Overall percentage	46.3	53.7	94.4	Overall percentage	48.1	51.9	92.6
DD		Predicted				Predicted		
	Observed	High (152 µm)	Low (≤ 152 µm)	Percent correct	Observed	High (152 µm)	Low (≤ 152 µm)	Percent correct
	High (> 152 µm)	25	1	96.2	High (> 152 µm)	25	1	96.2
	Low (≤ 152 µm)	2	26	92.9	Low (≤ 152 µm)	3	25	89.3
	Overall percentage	50	50	94.5	Overall percentage	51.9	48.1	92.6
WWR		Predicted				Predicted		
	Observed	High (> 0.079)	Low (≤ 0.079)	Percent correct	Observed	High (> 0.079)	Low (≤ 0.079)	Percent correct
	High (> 0.079)	24	2	92.30	High (> 0.079)	24	2	92.3
	Low (≤ 0.079)	3	25	89.28	Low (≤ 0.079)	4	24	85.7
	Overall percentage	50	50	90.80	Overall percentage	51.9	48.1	88.9

Conclusions

In this paper, two data mining-based classification techniques, i.e. CART and CHAID algorithms, are employed to generate the corresponding decision trees for analyzing a past WEDM experimental dataset. It is observed that in the WEDM process, T_{on} is the most important input parameter affecting almost all the responses, followed by T_{off} and I_P. On the other hand, WF and WT contribute least towards attainment of the target response values.

During T_{on}, actual machining in the WEDM process usually takes place. An increment in T_{on} causes the machining process to be faster with higher MRR and poor quality of the machined surface.

On the other hand, during T_{off}, the dielectric fluid in the WEDM process is re-ionized. An insufficient T_{off} time may lead to erratic cycling, thereby slowing down the machining process. With increased T_{off}, MRR is gradually decreased, primarily due to reduced spark discharge energy. Higher I_P causes availability of larger discharge energy resulting in more material being removed from the workpiece surface. Thus, a mix of higher T_{on} with lower T_{off} results in more sparking time, thus leading to increased MRR which is of primary importance in any of the machining processes.

The induction rules extracted from the decision trees lead to the following conclusions:

1. For achieving higher MR, high values of T_{on} and I_P, and low values of T_{off} and SV need to be set.
2. To attain lower SR of the machined components, low or medium T_{on}, high T_{off}, low or medium I_P and high SV are recommended. WF and WT play no significant roles on SR.
3. A mix of low or medium T_{on}, long T_{off}, small I_P and high SV is responsible for attainment of lower DD.
4. Low T_{off}, high or medium T_{on}, high I_P and low or medium SV are accountable for lower WWR.
5. The CART algorithm supersedes CHAID algorithm with respect to both overall classification accuracy and prediction risk. But, both these algorithms can almost perfectly envisage high and low values of all the considered responses for the WEDM process.

Thus, these data mining tools can be effectively applied to all the traditional and non-traditional machining processes to investigate the contributions of varied input parameters on the responses and identify the most suitable parametric combinations for exploring their fullest machining potential. But, the developed decision trees are highly unstable compared to other decision predictors as a small variation in data may result in a major change in the structure of the decision trees, conveying different pictures from the expected ones.

References

Agarwal S., Dhangde S., Chakraborty S. (2019). *Development of association rules to study the parametric influences in non-traditional machining processes.* Sadhana, 44, 230.

Arikatla S.P., Mannan K.T., Krishnaiah A. (2017). *Parametric optimization in wire electrical discharge machining of titanium alloy using response surface methodology.* Materials Today: Proceedings, 4,

Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1993). *Classification and Regression Tree.* New York: Chapman and Hall.

Dabade U.A., Karidkar S.S. (2016). *Analysis of response variables in WEDM of Inconel 718 using Taguchi technique.* Procedia CIRP, 41, 886–891.

Devarajaiah D., Muthumari C. (2018). *Evaluation of power consumption and MRR in WEDM of Ti-6Al-4V alloy and its simultaneous optimization for sustainable production.* Journal of the Brazilian Society of Mechanical Sciences and Engineering, 40, 8, 1–18.

Han J., Kamber M., Pei J. (2012). *Data Mining Concepts and Techniques.* USA: Elsevier Inc.

Ho K.H., Newman S.T., Rahimifard S., Allen R.D. (2004). *State of the art in wire electrical discharge machining (WEDM).* International Journal of Machine Tools & Manufacture, 44, 12-13, 1247–1259.

Kass G.V. (1980). *An exploratory technique for investigating large quantities of categorical data.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 29, 2, 119–127.

Kumar A., Kumar V., Kumar J. (2013). *Multi-response optimization of process parameters based on response surface methodology for pure titanium using WEDM process.* International Journal of Advanced Manufacturing Technology, 68, 9–12, 2645–2668.

Lal S., Kumar S., Khan Z.A., Siddiquee, A.N. (2015). *Optimization of wire electrical discharge machining process parameters on material removal rate for Al7075/SiC/Al_2O_3 hybrid composite.* Proceedings of the Institution of Mechanical Engineers Part B: Journal of Engineering Manufacture, 229, 5, 802–812.

Lusi N., Muzaka K., Soepangkat B.O.P. (2016). *Parametric optimization of wire electrical discharge machining process on AISI H13 tool steel using weighted* *principal component analysis (WPCA) and Taguchi method*. ARPN Journal of Engineering and Applied Sciences, 11, 2, 945–951.

Mandal A., Dixit A.R. (2014). *State of art in wire electrical discharge machining process and performance*. International Journal of Machining and Machinability of Materials, 16, 1, 1–21.

Manjaiah M., Laubscher R.F., Kumar A., Basavarajappa S. (2016). *Parametric optimization of MRR and surface roughness in wire electro discharge machining (WEDM) of D2 steel using Taguchi-based utility approach*. International Journal of Mechanical and Materials Engineering, 11, 7, 2–9.

Nayak B.B., Abhishek K., Mahapatra S.S. (2018). *Parametric appraisal of WEDM taper cutting process using maximum deviation method*. Materials Today: Proceedings, 5, 11601–11607.

Patel V.D., Vaghmare R.V. (2013). *A review of recent work in wire electrical discharge machining (WEDM)*. International Journal of Engineering Research and Applications, 3, 3, 805–816.

Ramanan G., Elangovan R. (2018). *Parametric optimization of wire cut electrical discharge machining on Al-9% PAC composites using desirability approach*. International Journal of Vehicle Structures and Systems, 10, 6, 467–470.

Ramani J., Dandge S., Chakraborty S. (2020). *Machinability study of plain carbon steels using data mining technique*. AIP Conference Proceedings, 2273, 050005.

Sarker B., Chakraborty S. (2021). *Structural equation modeling-based performance estimation and parametric analysis of wire electrical discharge machining processes*. Sadhana, 46, 1–14.

Srinivasarao G., Suneel D. (2018). *Parametric optimization of WEDM on α-β titanium alloy using desirability approach*. Materials Today: Proceedings, 5, 7937–7946.

Tan P.N., Steinbach M., Kumar V. (2006). *Introduction to Data Mining Instructor's Solution Manual*. USA: Pearson Addison Wesley.

Vignesh M., Ramanujam R. (2018). *Response optimization in wire electrical discharge machining of AISI tool steel using Taguchi GRA approach*. International Journal of Machining and Machinability of Materials, 20, 5, 474–495.

S S Dandge

Shankar Chakraborty 21/04/25