Improved machine learning approaches designed and validated in different engineering application fields

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Maschinenbau und Verfahrenstechnik
der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades
einer
Doktorin der Ingenieurwissenschaften
Dr.-Ing.
genehmigte Dissertation

von

Xiao Wei
aus
Liaoning, China

Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker
Prof. Dr. rer. nat. Robert Martin

Tag der mündlichen Prüfung: 04.05.2023

To my parents

Acknowledgements

This thesis is the result of the research work I carried out at the Chair of Dynamics and Control (SRS) at the University of Duisburg-Essen during 2019 to 2023. I would like to thank all the people who have supported and guided me during this important period in my life.

First and foremost, I would like to express my special appreciation and thanks to my supervisor, Univ.-Prof. Dr.-Ing. Dirk Söffker, for offering the opportunity to be his student, for the continuous guidance and great support he has given me during my Ph.D. study. Without his valuable advices and constant feedback this work would not have been achievable.

I am also grateful to my second supervisor, Prof. Dr. rer. nat. Robert Martin, for his valuable feedback, scientific questions, and helpful advices to improve this work.

I would like to thank all my colleagues at the Chair of Dynamics and Control (SRS) for their kind support, encouragement, and cooperation.

I also thanks my friends in Germany who help me to overcome the difficulties in study and living at the beginning I came here. Besides, thanks a lot of my friends in China who help me wit the things in life.

A very special word of thanks goes for my mama and my sister who always encourage and believe in me.

Most importantly, I would like to thank my papa, who encouraged me to explore the world with a grateful heart.

Mülheim, May 2023 Xiao Wei

Kurzfassung

Maschinelles Lernen (ML) bietet Maschinen die Fähigkeit, automatisch aus Daten und vergangenen Erfahrungen zu lernen, um Muster zu erkennen und mit minimalem menschlichem Eingreifen Vorhersagen für neue Daten zu treffen. Im Vergleich zu anderen statistischen Technologien kann ML die Berechnung kontinuierlich verbessern, Entscheidungen automatisch treffen und Trends und Muster automatisch erkennen. Basierend auf der Tatsache, dass mehr Datensätze für die Öffentlichkeit zugänglich sind, verbesserte Computerberechnungsfähigkeiten und eine Beschleunigung der Internetgeschwindigkeit, haben sich maschinelle Lernalgorithmen in den letzten Jahrzehnten schnell entwickelt. In den letzten Jahren wurden zahlreiche neue ML-Ansätze entwickelt und aktualisiert. Die meisten dieser entworfenen Ansätze zielen jedoch auf einen bestimmten Datensatz ab, daher ist die Verallgemeinerung dieser Ansätze nicht verifiziert. Neben dem spezifisch entworfenen Ansatz für einen Datensatz sind viele vortrainierte Modelle (die durch eine große Anzahl von Daten vorgeschlagen werden) öffentlich zugänglich. Die meisten dieser vortrainierten ML-Modelle betreffen jedoch auch ein spezielles Gebiet, und ihre Leistung ist nicht ideal, wenn sie auf Datensätze in anderen Gebieten angewendet werden. Folglich sind verbesserte ML-Ansätze erforderlich, die in verschiedenen Bereichen angewendet werden können.

Um ML-Ansätze zu entwerfen, die in verschiedenen Bereichen verfügbar sind, werden vier Datensätze aus drei Anwendungsbereichen verwendet, um die in dieser Studie entworfenen Ansätze zu verifizieren. Der Datensatz zur inneren Sprache (IS) gehört zum Bereich Biologie. Die Frequenzbänder des IS-Datensatzes sind sehr niedrig (von 0,5 Hz bis 100 Hz). In diesem Datensatz werden während des Experiments verschiedene Zeichen durch innere Sprache ausgelöst. Elektroenzephalogramm (EEG)-Signale, die während des IS-Verfahrens erfasst werden, und Daten werden analysiert und klassifiziert. Der Lagerdatensatz der Case Western Reserve University (CWRU) wird von der mechanischen Schule in CWRU durchgeführt. Es wird häufig zur Verifizierung von Lagerdiagnoseansätzen verwendet. Vibrationssignale werden erhalten, wenn während des Experiments verschiedene fehlerhafte Lager verwendet werden. Vibrationssignale haben ein breites Frequenzband von Dutzenden Hz bis Tausend Hz. Ein weiteres Anwendungsgebiet, das in dieser Studie untersucht wird, sind Metallbearbeitungsflüssigkeiten (KSS). Akustische Emission (AE)-Signale, wenn eine Vielzahl von MWF während des Gewindeformprozesses angewendet werden, werden analysiert. Schallemissionssignale sind transiente Spannungswellen, die durch die schnelle Freisetzung von Energie aus festen Quellen mit Frequenzbändern bis zu Millionen Hz erzeugt werden. Je nach MWF-Typ werden zwei Experimente durchgeführt und relevante AE-Signale erfasst. Im ersten Versuch (Datensatz MWF19) werden nur Kühlschmierstoff (KSS) auf Emulsionsbasis eingesetzt. Im zweiten Experiment (Datensatz MWF16) werden sowohl emulsionsbasierte als auch ölbasierte KSS verwendet.

Entsprechend den verschiedenen Highlights und Schritten im maschinellen Lernen wer-

den in dieser Studie drei Kategorien mit 5 Unteransätzen entworfen. Ansatz 1 enthält einige Schritte zur Datenverarbeitung - Datenauswahl, Segmentierung und Klassifizierung. Bei diesem Ansatz werden Signale im Zeitbereich analysiert. Die Konzentration von Ansatz 1 liegt auf der Struktur des Klassifikators (Convolution Neural Network (CNN)) und Hyperparametern, die sich auf das Design und die Abstimmung von Trainingsalgorithmen beziehen. Das neuronale Faltungsnetzwerk wird entsprechend der Probenlänge und Funktionalität jeder Schicht entworfen. Neben der CNN-Struktur und Hyperparametern in CNN wird ein neues Datenverarbeitungsverfahren für die Differenzierung verschiedener Phasen in einer Messung entwickelt. Da die Grenzen zwischen verschiedenen Phasen in MWF-Datensätzen im Zeitbereich nicht klar sind, wird die kontinuierliche Wavelet-Transformation (CWT) als Werkzeug angewendet, um Grenzen zwischen verschiedenen Teilen in einer Messung zu finden. Darüber hinaus wird die Länge des Segments durch die Rotationsgeschwindigkeit des Werkzeugs im Datensegmentierungsschritt definiert.

Im Vergleich zu Ansatz 1 sind in Ansatz 2 mehr Datenverarbeitungsmethoden enthalten. Obwohl in diesem Ansatz dieselbe Datenauswahlmethode wie in Ansatz 1 verwendet wird, werden neben der Datenauswahl und -segmentierung in Ansatz 1 Daten vom Zeitbereich in den Zeitfrequenzbereich transformiert durch Kurzzeit-Fourier-Transformation und Spektrogramme erhalten. Segmente und Spektrogramme werden normalisiert, bevor sie in den Klassifikator eingegeben werden. Anders als die Messung in Segmente fester Größe gemäß der Werkzeuggeschwindigkeit segmentiert wird, ist die Segmentlänge bei diesem Ansatz nicht identisch, sie werden als ein einstellbarer Parameter betrachtet. Außerdem sind Überschneidungen zwischen verschiedenen Klassen nicht identisch, wenn die Probennummern in jeder Klasse nicht gleich sind. Zusätzlich werden Parameter in der Datenverarbeitung und Hyperparameter in CNN zusammen und automatisch in einem Schritt optimiert. Darüber hinaus wird, basierend auf Ansatz 2, ein Transfer Learning (TL)-Ansatz (Ansatz 2.0) zwischen MWF19 und MWF16 erhoben. Parameter in der Datenverarbeitung und Hyperparameter in CNN, die von MWF19 trainiert wurden, werden in Ansatz 2.0 in MWF16 transformiert.

Anders als in Ansatz 1 und 2 eine Methode oder ein Algorithmus in einem Schritt angewendet wird, werden in Ansatz 3 verschiedene Datenverarbeitungsmethoden und ML-Algorithmen in jedem Schritt ausprobiert. Aus dieser Sicht ist Ansatz 3 eine Integration verschiedener Ansätze anstelle eines einzelnen Ansatzes . Bei Ansatz 1 und Ansatz 2 werden feste Daten verwendet, im Gegensatz dazu werden bei diesem Ansatz verschiedene Teildaten ausprobiert. Außerdem werden in jedem Schritt viele ähnliche, aber nicht identische Funktionsdatenverarbeitungsverfahren und ML-Algorithmen verwendet. Ansatz 3 wird anhand der Destinationsdifferenz in Supervised Learning (Ansatz 3.1) und Unsupervised Learning (Ansatz 3.2) unterteilt. Obwohl der Prozess von Ansatz 3.1 und 3.2 derselbe ist, unterscheiden sich die Methoden, die in jedem Schritt von Ansatz 3.1 und 3.2 verwendet werden. In Ansatz 3.1 werden verschiedene Methoden zur Aufteilung von Training-Testdaten diskutiert. Außerdem werden im Datenverarbeitungsschritt eine Rohmessung, ein Savitzky-Golay (SG)-Filter und eine empirische Modenzerlegung (EMD) angewendet. Als Klassifikator werden lineare, polynomiale und gaußsche Kernel SVM verwendet. In Ansatz 3.2 werden STFT, CWT und Hilbert-Huang-Transformation (HHT) für die Signaltransformation vom Zeitbereich in den Zeit-Frequenz-Bereich verwendet. Als Merkmalsextraktionsverfahren wird alternativ Autoencoder eingesetzt. K-Mittelwert und Gaußsches Mischungsmodell (GMM) werden für das Feature-Clustering verwendet.

Ansatz 1, 2 und 3.2 werden auf CWRU- und MWF-Datensätze angewendet. Da sich

der IS-Datensatz stark von anderen Datensätzen unterscheidet, ist Ansatz 3.1 spezifisch für ihn. Außerdem wird in Anbetracht der Ähnlichkeit zwischen MWF19 und MWF16 der Ansatz 2.0 zwischen ihnen angewendet. Es wird die Abhängigkeit der Wahl von Trainings- und Testdaten auf das Ergebnis diskutiert, die im Fall des IS-Datensatzes erheblich sind. Die besten Ergebnisse des CWRU-Lagerdatensatzes aus dem überwachten Lernansatz (Ansatz 2) sind der F-Score und die Genauigkeit beträgt 100 % für 29 Lagerzustände, was bedeutet, dass alle Lagerzustände im CWRU-Lagerdatensatz perfekt klassifiziert werden können. Die besten Ergebnisse des CWRU-Lagerdatensatzes aus dem unüberwachten Lernansatz (Ansatz 3.2) sind, dass alle Metriken 1,0 bei der Unterscheidung zwischen fehlerfreien und fehlerhaften Lagern sind. Für den MWF19-Datensatz stammen die besten Klassifizierungsergebnisse aus Ansatz 2 – F-Score und Genauigkeit sind 98,61 % und 98,58 %. Außerdem werden Referenz- und andere Fluide vollständig durch Ansatz 3.2 unterschieden. Für den MWF16-Datensatz erreichen die besten Klassifikationsergebnisse eine Genauigkeit von 98,11 % von Ansatz 1. Obwohl die Clustering-Ergebnisse für MWF16 noch offen sind, können noch einige Schlussfolgerungen aus dem Berechnungsprozess gezogen werden.

Abstract

Machine learning (ML) provides machines the ability to automatically learn from data and past experience to identify patterns and make predictions for new data with minimal human intervention. Comparing with other statistical technologies, ML can improve calculation continuously, making decisions automatically, and identifying trends and patterns automatically. Based on the fact more datasets open to public, improved computer calculation capability, and internet speed acceleration, machine learning algorithms develop fast in the past decades. Large number of new ML approaches are raised and updated in the last years. However, most of these designed approaches target on one specific dataset, therefore, generalization of these approaches are not verified. Besides on the specific-designed approach to one dataset, a great deal of pre-trained models (which are proposed by a large number of data) are open to public. However, most of these pre-trained ML models are also concerning to one special field and their performance is not ideal when applied to datasets in other fields. Consequently, improved ML approaches that can be applied in various fields are required.

To design ML approaches available in various fields, four datasets from three application fields are employed to verify approaches designed in this study. Inner speech (IS) dataset belongs to biology field. The frequency bands of IS dataset are very low (from 0.5 Hz to 100 Hz). In this dataset, various signs are acted by inner speech during the experiment. Electroencephalogram (EEG) signals acquired during IS procedure and data are analyzed and classified. Case Western Reserve University (CWRU) bearing dataset is conducted by the mechanical school in CWRU. It is often used for bearing diagnosis approaches verification. Vibration signals are obtained when various faulty bearings are employed during the experiment. Vibration signals has a wide frequency bands from dozen Hz to thousand Hz. Another application field employed in this study is metalworking fluids (MWF). Acoustic Emission (AE) signals when a variety of MWF are applied during thread forming process are analyzed. Acoustic Emission signals are transient stress waves generated by the rapid release of energy from solid sources with frequency bands up to million Hz. According to the MWF types, two experiments are conducted and relevant AE signals are acquired. In the first experiment (dataset MWF19) only emulsion-based MWF are applied. Both emulsion-based and oil-based MWF are used in the second experiment (dataset MWF16).

According to the different highlights and steps in machine learning, three categories with 5 sub approaches are designed in this study. Approach 1 contains few steps on data processing - data selection, segmentation, and classification. In this approach, signals are analyzed in time domain. Concentration of approach 1 is on classifier (convolution neural network (CNN)) structure and hyperparameters referring to training algorithms designing and tuning. Convolution neural network is designed according to samples length and functionality of each layer. Beside the CNN structure and hyperparameters in CNN, a new

data processing method is designed for various phases differentiation in one measurement. As the boundaries among different phases are not clear in time domain in MWF datasets, continuous wavelet transform (CWT) is applied as a tool to find boundaries among different parts in one measurement. Furthermore, segment's length is defined by rotating tool speed in the data segmentation step.

Comparing with approach 1, more data processing methods are included in approach 2. Although the same data selection method is used in this approach as approach 1, besides data selection and segmentation in approach 1, data are transformed from time domain to time-frequency domain by Short-time Fourier Transform (STFT) and spectrograms are obtained. Segments and spectrograms are normalized before they are put into the classifier. Unlike measurement are segmented into fixed size segments according to tool speed, segment length is not identical in this approach, they are considered as one adjustable parameter. Besides, overlap among different class are not identical when samples numbers in each class are not the same. Additionally, parameters in data processing and hyperparameters in CNN are optimized together and automatically in one step. Furthermore, based on approach 2, a transfer learning (TL) approach (approach 2.0) is raised between MWF19 and MWF16. Parameters in data processing and hyperparameters in CNN trained from MWF19 are transformed to MWF16 in approach 2.0.

Unlike one method or algorithm is applied in one step in approach 1 and 2, varied data processing methods and ML algorithms in each step are tried in approach 3. From this point of view, approach 3 is a integration of various approaches instead of single approach. Fixed data are used in approach 1 and approach 2, on the contrary, various parts data are tried in this approach. Besides, many similar but not identical function data processing methods and ML algorithms are employed in each step. Based on the destination difference, approach 3 is divided into supervised learning (approach 3.1) and unsupervised learning (approach 3.2). Although the process of approach 3.1 and 3.2 is the same, methods used in each step in approach 3.1 and 3.2 are varied. In approach 3.1, different training-test data split ways are tried. Besides, raw measurement, Savitzky–Golay (SG) filter, and empirical mode decomposition (EMD) are applied in data processing step. Linear, polynomial, and Gaussian kernels SVM are employed as classifier. In approach 3.2, STFT, CWT, and Hilbert–Huang transform (HHT) are used for signals transformation from time domain to time-frequency domain. Autoencoder as an alternative is employed as feature extraction method. K-mean and Gaussian mixture model (GMM) are used for features clustering.

Approach 1, 2, and 3.2 are applied to CWRU and MWF datasets. As the IS dataset is very different from other datasets, so approach 3.1 is specific to it. Besides, considering the similarity between MWF19 and MWF16, the approach 2.0 is applied between them. For IS dataset, when training-test data are split in different ways, the results are significantly different. Best results of CWRU beaing dataset from supervised learning approach (approach 2) are F-score and accuracy is 100% for 29 bearing states, which denotes that all bearing states in CWRU bearing dataset can be classified perfectly. Best results of CWRU bearing dataset from unsupervised learning approach (approach 3.2) is all metrics are 1.0 in fault-free and faulty bearing distinction. For MWF19 dataset, the best classification results are from approach 2 - F-score and accuracy are 98.61 % and 98.58 %. Beside, reference and other fluid are totally distinguished by approach 3.2. For MWF16 dataset, the best classification results arrive to accuracy is 98.11 % from approach 1. Although clustering results are still open for MWF16, some conclusions can still be drawn from calculation process.

Contents

List of Abbreviations

AE	Acoustic Emission
AI	Artificial intelligence
ARI	Adjusted rand index
ANN	ArtificIal neural network
B	Ball
BCI	Brain-computer interface
BLSTM	Bidirectional long short-term memory
CML	Conventional machine learning
CNN	Convolution neural network
CWRU	Case Western Reserve University
CWT	Continuous wavelet transform
DBN	Deep belief network
DE	Drive end
DL	Deep learning
DT	Decision tree
DWT	Discrete wavelet transform
ECoG	Electrocorticography
EEG	Electroencephalogram
EMD	Empirical mode decomposition
FC	Fully connected
FE	Fan end
fMRI	functional magnetic resonance imaging

fNIRS	functional near infrared spectroscopy
FT	Fourier transform
GAN	Generative adversarial network
GD	Gaussian distribution
GMM	Gaussian mixture model
GPR	Gaussian process regression
GUI	Graphic user interface
HHT	Hilbert-Huang transform
HMM	Hidden mixture model
IMF	Intrinsic mode function
IR	Inner raceway
IS	Inner speech
K-NN	K-nearest neighbor
LDA	Linear discriminant analysis
MEG	Magnetoencephalography
ML	Machine learning
MNI	Normalized mutual information
MWF	Metalworking fluid
NB	Naive Bayes
OR	Outer raceway
RF	Random forest
RNN	Recurrent neural network
RI	Rand index
RVM	Relevance vector machine
SD	Standard deviation
sEEG	stereotactic electroencephalogram
SG filter	Savitzky-Golay (SG) filter

SLR	Sparse logistic regression
SOM	Self-organization maps
STFT	Short-time Fourier transform
SVM	Support vector machine
TFA	Time-frequency analysis
TL	Transfer learning
WT	Wavelet transform

List of Mathematical Notations

$f(t)$	Signal
b	Desired location on time scale
a	Dilation parameter on time and frequency scale
$\phi(t)$	Short-time Fourier transform window function
$\Psi(t)$	Wavelet transform window function
$H[g(t)]$	Hilbert transform
p	Polynomial order
δ	Kernel scale
ω_i	Mixture weights
μ_i	Mean vector
σ_i	Covariance matrix
$g(X\|\mu, \sigma_i)$	Component Gaussian densities
$I(Y;C)$	Mutual information between Y and C

List of Figures

List of Tables

1 Introduction

Artificial intelligence (AI) which imitate human abilities brings a promise of genuine human-to-machine interaction. As a subset of AI, machine learning (ML) allows machines to learn from data without being explicitly programmed [JK20]. Machine learning enables a computer system to make predictions or take some decisions using historical data without being explicitly programmed [OCBR18]. A large amount of structured and semi-structured data are applied so that a machine learning model can generate accurate result or give predictions [GSA+18]. Machine learning has a large number of advantages. Firstly, ML is capable of improvement continuously, when new data are provided, model's accuracy and efficiency improve with subsequent training [DSD+21]. Secondly, ML can fulfill various decision-making tasks automatically such as data classification, regression, and clustering [LIS16]. Lastly, ML has the ability of identifying trends and patterns. Because of the advantages, ML has wide application in the fields of image and speech recognition, traffic prediction, product recommendation, self-driving cars, e-mail spam and malware filtering, virtual personal assistant, stock market trading, medical diagnosis, fault diagnosis and prognosis, and automatic language translation.

According to different ways in data processing, ML can be divided into conventional machine learning (CML) and deep learning (DL). Conventional machine learning techniques have limited capability of processing data in their original form [CS18]. These methods required considerable understanding and expertise for data features [LSSG17]. In other words, before original data are put into the CML, data need to be processed and features should be extracted. A large number of data processing methods are available such as: data selection, normalization, segmentation, time-frequency transformation, and augmentation. On the contrary, DL eliminates some data pre-processing steps that is typically involved in CML and it can ingest and process unstructured data. No matter CML or DL, machine learning algorithms generate models on training data. When new input data are introduced to the trained model, ML uses the developed model to make a prediction for new input data. The performance of the models are evaluated by metrics. Based on the task, a large number of metrics are available. The models are either deployed or trained repeatedly until the desired metrics values are achieved. According to [ELM20], the process of ML is divided into 3 steps: decision process (based on some input data labeled or unlabeled, the algorithm produce a model from the input data); error function (evaluate the performance of trained models); model optimization (update weights autonomously until a metric threshold acquired).

Machine learning algorithms are categorized into three types: supervised learning, unsupervised learning, and reinforcement learning according to research destination [SK17]. Supervised learning and unsupervised learning approaches are used in most cases. Supervised learning algorithms are trained on data with labels and usually for the task of classification and regression [MJJL08]. The commonly used supervised learning methods

are: Naive Bayes (NB), linear regression, logistic regression, support vector machine (SVM), K-nearest neighbor (K-NN), random forest (RF), decision trees (DT), Gaussian process regression (GPR), convolutional neural network (CNN), generative adversarial networks (GAN), and self-organization maps (SOM). Unsupervised learning are able to work with data without labels [WLZ+16]. The commonly used unsupervised learning methods are: K-means, hierarchical cluster, Gaussian mixture model (GMM), fuzzy C-means, autoencoder, and deep belief network (DBN).

1.1 Motivation and problem statement

Data are the base stone for ML, models can not be trained without data. To facilitate the training of ML model, many databases are open to the public such as IRIS, MINIST, Boston Housing, Fake News Detection, and SOCR datasets. These benchmark datasets are used for training new models and compare trained models to other approaches. However, most of these datasets are specific to one field, for exaple, IRIS dataset is used for iris plants distinction while MINIST dataset is applied for handwriting distinguish. In addition to public datasets, a large number of pre-trained models are open to public to reduce time consuming in model training processes, such as $LeNet, AlexNet, GoogLeNet, VGGNet, ResNet$. However, these pre-trained DL models performance are not ideal when they are used for datasets from other fields. As a results, a large number of new approaches are designed and proposed. Most of these designed approaches target on one specific dataset, therefore, generalization of these approaches are not verified. In this study, different approaches that could be used for more datasets are designed and proposed. The performance of these proposed approaches are validated on datasets from different fields. Four datasets in three fields from brainwaves, vibration, and Acoustic Emission (AE) signals are applied to verify proposed approaches.

Brainwaves are electrical impulses and they appear when individual's behavior, emotions, and thoughts are communicated between neurons [SLS17]. Brainwaves occur at various frequency bands from 0.5 Hz to more than 100 Hz. Inner speech (IS) is the silent expression of conscious thought to oneself in a coherent linguistic [VPB13]. When people suffer certain brain disorders, they cannot communicate normally. In this case, the best solution for them is inferring IS directly form brainwaves. Electroencephalogram (EEG) is one typical non-invasive methods to capture brainwaves [SI15]. In case that EEG signals during IS can be well identified, it will bring great convenience to individuals with brain disorders. Nowadays many contributions focus on the relationship between IS and EEG signals and some of them achieved good results from physical point of view, however, there are still many issues worthy of further research, especially classification of the EEG signals from different symbols, words, and signs are still not clearly. The disadvantages of these approaches are: firstly, most contributions apply their approach distinguishing EEG signals to the datasets from their own experiments, so their approaches generalization cannot be validated. Secondly, results from these approaches are improvable - most distinction accuracy among different words or vowels are under 50 % and stand deviation (SD) among different subjects are very high.

Rotating machinery plays an important role in industry from the last two centuries. Several surveys regarding the likelihood of induction machine failures conducted by the IEEE Industry Application Society (IEEE-IAS) and the Japan Electrical Manufacturers' Association (JEMA) reveal that bearing faults are the most common fault type and are

responsible for 30 to 40 % of all the machine failures [ZZWH20]. Consequently, bearing fault diagnosis and prognosis is of increasing interest and several benchmark datasets has been developed. Among these datasets, Case Western Reserve University (CWRU) bearing dataset is well accepted as benchmark used for bearing diagnosis and allows to compare approaches using statistical measures. Large number of new diagnosis and prognosis approaches are designed for rotating machinery in the past years. Although results from these approaches are good - most contributions arrive accuracy over 90 % and some even arrive 100 %, inadequacy of these approaches are still need to be explored. Firstly, selected data that used in relevant approaches are not clear in some contributions. While data are the most critical part in machine learning, results are not convincing when data used are not pointed out. Then, for some contributions, only part of the data are selected - data from drive side faulty bearing and baseline are used. Data from drive end faulty bearing and baseline are easier to differentiate, good results from easily distinguishable data can not demonstrate the robustness of these approaches. Furthermore, most study classify selected data into 4 or 10 classes. Four classes denoted as fault-free, fault at inner race, ball, and out race while ten classes denoted as fault-free, fault location in different parts (inner race, ball, out race) combing with fault size (7, 14, 21 mils). Fault size of 28 mils and different fault locations in out race can not be distinguished in four or ten classes. Besdies, in some publications only a small amount of information is referred on data processing, parameters, and hyperparameters optimization. If only parts of well-performed samples are taken into calculation, results are unconvincing. Lastly, accuracy is mostly applied to evaluate the performance of these approaches. When sample numbers in each class are not equal, accuracy is not suitable to evaluate approaches and other metrics are needed to judge approaches performance like F-score.

Metalworking fluid (MWF) plays a significant role in manufacturing processes as it cool and lubricate the contact zone between tool and workpiece to prevent tool wear and to ensure manufacturing of required geometries and surface qualities [BMHCH15]. Metalworking fluid and its additives mainly affect tool wear and workpiece surface roughness or make higher machining speeds possible to decrease manufacturing time and increase the output [LCN12]. Transient stress waves generated by the rapid release of energy from localized sources which is also name as AE during metalworking process [SW83]. Acoustic Emission can be detected by sensors attached to or embedded in the structure being monitored and it can be linked to the onset of new damage or to the progression of existing anomalies. To verify the relations between MWF and AE on tool and workpiece during thread forming, two experiments are conducted by the collaboration between the Chair of Dynamics and Control, University of Duisburg-Essen and Rhenus Lub GmbH & Co KG Mönchengladbach. Consequently, two datasets (MWF19 and MWF16) are acquired. Although less contributions employ them as benchmark as these two datasets are still not open to public, Mr. Wirtz and Miss Dermmerling apply them in their research. Approach applied by Mr. Wirtz and Miss Dermmerling is identical to these two datasets - data in both two datasets are distinguished by the combination of continuous wavelet transform (CWT) and K-means. Besides, parts of data are applied by these two authors. Results from their approach is not ideal and can be improved. Furthermore, performance of relevant approaches are just shown in photos, no detailed values are presented. So these approach cannot be evaluated comprehensively. More metrics are needed to evaluate proposed models.

1.2 Thesis organization

This thesis is organized as follows:

In the second Chapter, theoretical background are introduced briefly. Data need to be processed and their features should be extracted before ML. Various data processing methods are applied in this study, such as selection, segmentation, Savitzky-Golay (SG) filter. In addition to data filter, signals are often analyzed by time-frequency analysis (TFA) technology such as: short-time Fourier transform (STFT), wavelet transform (WT), empirical mode decomposition (EMD), and Hilbert-Huang transform (HHT). Furthermore, data are normalized by Z-score normalization and Min-Max normalization technology. Besides data processing methods, ML algorithms are also described. Machine learning algorithms employed in this study involve: convolution neural network (CNN), support vector machine (SVM), autoencoder, K-means, Gaussian mixture models, and transfer learning (TL). To optimize parameters in data processing and hyperparamters in ML, both exhaustive sweep and Bayesian optimization are used. Various metrics used for supervised learning and unsupervised learning separately in this study. Accuracy, F-score, and SD are applied for supervised learning. Purity, rand index (RI), adjusted rand index (RAI), normalized mutual information (NMI), and F-measure are applied for unsupervised learning.

In chapter 3, datasets applied in this study and their experiments are presented. To verify the proposed approaches, three kinds signals are employed: EEG signals from IS, vibration signals from CWRU bearing dataset, and AE signals from various MWF during thread forming process. Electroencephalogram signals are brainwaves which have low frequency (from 0 to hundred Hz), vibration signals belong to middle frequency rang (from dozen to thousand Hz), and AE signals have very high frequency (from thousand to million Hz). Besides, these three kinds of signals belong to different application fields: EEG signals belong to biology field, CWRU bearing dataset belongs to mechanical engineering field, and AE signals from varied MWF belong to cross disciplinary. The way acquiring data in these datasets are extremely differential. Test rig and experimental procedure of these three kind of signals are explained in detail in this chapter.

Basic information on ML and state-of-art approaches applied in these three kind of datasets are summarized in chapter 4. The results relevant to these approaches are exhibited. In the first section, general information on ML are presented. Experiments and signals referring to IS, approaches applied in IS distinction are summed up and their results are reviewed in the second section. As CWRU bearing is a benchmark, a large numbers of approaches are verified by it. These approaches are summarized and their results are compared in the third section. Lastly, approaches applied to the MWF datasets and their results are presented in this chapter.

In chapter 5, the proposed approaches are presented. The first approach is mainly focus on CNN hyperparameters optimization - structure and hyperparameters referring to training algorithms are tuned according to each datasets. Besides, a new method is raised for data selection in approach 1. More data processing methods (data selection, segmentation, transformation, and normalization) are applied in the second approach. Besides, parameters in data processing steps and hyperparameters in CNN are optimized in one step and a loop is established among data segmentation and classification. Based on the approach 2, TL (approach 2.0) is applied between two MWF datasets. The approach 3 is a combination of different data selection, data processing and ML methods.

Both supervised (approach 3.1) and unsupervised learning approaches (approach 3.2) are designed. Flowchart, highlight, and innovation of each approach are presented in this chapter.

When proposed approaches are validated by four datasets, their results are shown in chapter 6. As five sub approaches are designed, numbers combination can be gotten when they are applied to these four datasets. To express the results clearer, results are shown according to dataset. For IS dataset, results from approach 3.1 are presented. For CWRU dataset, results from approach 1, 2, and 3.2 are shown. For MWF19 datasets, results from approach 1, 2, and 3.2 are presented. Results from approach 1, 2.0, and 3.2 to MWF16 dataset are also shown in this chapter. In addition, results from proposed approaches are also compared with other literature. By results comparison, the advantages of proposed approaches are emphasized.

Summary, conclusions, and outlook are parts of the last chapter. Contents of this study are summarized in the first section. Conclusions from calculation process, results and results comparison are drawn in the second section. Expectations and ideas for the future work are referred in the third section of this chapter.

2 Theoretical background

Artificial intelligence is a big concept to create intelligent machines that can simulate human thinking capability and behavior [AN21]. Nowadays it has become an all-encompassing term and has wide applications for complex tasks that used to require human input such as communicating with customers online or playing chess. As a subset of AI, ML allows machines to learn form data without being programmed explicitly [Mah20]. Machine learning is also a branch of computer sciences which focus on the use of data and algorithms to imitate the way that humans learn, gradually improving accuracy [AI22]. Machine learning could be divided into CML, DL, and TL depending on how the data is processed. Conventional machine learning techniques have limited in capability of processing the data in their original form. These methods require considerable understanding and expertise for representation i.e. selection of features is required carefully [CS18]. Conventional machine learning refers to a set of algorithms that permits a system to be input with dataset and it automatically realizes the representations required for decision making i.e. detection or classification [MBD+90]. On the contrary, DL has the ability to automatically extract, analyze, and understand the useful information from the raw data [KAF+08]. Transfer learning is an advanced machine learning method that a model developed for a task is reused as the starting point for a model on a second task [TS09]. According to the task difference, ML could be classified into supervised learning which is for the task of classification and regression and unsupervised learning which is for the task of clustering. With the development of computer technology and statistical methods, various kinds of ML algorithms and methods are available. Only those ML algorithms that relevant to this study will be introduced here.

Data processing has significant influence on CML, as it is usually treated as classifier without feature extraction function. Although DL has the functionality of features extracting, data processing are also needed to prompt data distinguishing results. Broadly speaking, data processing includes data collection, preparation, input, mining, interpretation, and storage [BF00]. However, in this study, as data have been collected during experiments, prepared and stored in electrical equipment, here data processing is specific to data mining and interpretation. A large number of data processing methods are on the list, only those methods that referring to this study are introduced here.

The contents, figures, and tables presented in this chapter are modified after previous publications [WS20] and [WJDS22]. Parts of the contents are prepared and submitted for publications of [WJS22], [WSS22], and [WLS22].

2.1 Data processing methods

Data selection, filter, transformation, and normalization methods are often employed in data processing. While there are often some irrelevant data (noise) in datasets from experiments

and manufacturing process. Data selection is of great significance as it remove irrelevant part data and keep the most useful parts data. Data filter has a significant impact on data processing as it remove outliers in signals [Lee21]. Data segmentation divide signals into more characteristic segments and helps to add sample number. Time-frequency analysis (TFA) help to draw signals time-frequency information from time domain. Normalization technologies remove the outliers and adjust data into a stable range for the convenience of calculation.

2.1.1 Data selection

Unlike benchmark datasets that only contains concerned data, most original datasets often contain a lot of irrelevant data in datasets from experiments and manufacturing process. Data selection aims at choosing proper data that should be archived in calculation. A number of issues should be considered of during data selection [Joc21]. These issues include:

i) Appropriate phases and sources of data which permit investigators to adequately answer the stated research questions,

ii) Suitable procedures in order to obtain a representative sample.

Generally, data applied or discarded should be defined by researchers under certain circumstances. While data selection procedures affect result greatly, they need to be considered thoroughly. Determining appropriate data is discipline-specific and is primarily driven by the goal of the investigation, existing literature, and accessibility to data sources. Before data selection, destination of research, scope of investigation, literature review, and data type should be in consideration. Data selection approaches and models are reviewed in [OC03].

2.1.2 Date filter

Measurement Segmentation

Measurement segmentation denote as divide whole measurement into various segments. Measurement segmentation has three functionalities according to [Joc21]. First of all, the complete measurement will be subtracted by irrelevant data segments. For example, this helps to find sections of a measurement that contains useful data. The second function is to multiply the number of data samples by dividing one measurement into multiple segments. The third functionality is to make sure that each sample which will be used for training and validation of the model has the same length. Segment length also named as segment size, it depends on measurement physical situation. In the process of segmentation, to maximize the number of samples, the segment size will be set to the lowest possible size that covers the smallest periodical event which contains all measurement information. For example, for rotating machinery, one segment should contains at least data in one round. Additionally, overlapping among segments is also used to multiply samples number. Besides, overlap among segments help to keep the important information than happens in-between segments.

Savitzky-Golay filter

Since it's introduction more than half a century ago, Savitzky-Golay (SG) filter has been popular in many fields of data processing: ranging from spectra in analytical chemistry,

via geography sciences to medicine [SRD22]. Both Savitzky and Golay pointed that fitting a polynomial to a set of input samples, then evaluate the resulting polynomial at a single point within the approximation interval which is equivalent to discrete convolution with a fixed impulse response [Sch11]. The low pass filter obtained by this method are SG filter. Savitzky-Golay smoothing filter is typically used to 'smooth out' a noisy signal whose frequency span (without noise) is large. It is also called digital smoothing polynomial filter or least-squares smoothing filter [RM13].

According to [Sch11], basic idea of least-squares polynomial smoothing is as following: $x[n]$ are samples of a signal as solid dots, considering to the moment the group of 2M+1 samples centered at $n = 0$, the coefficient of a polynomial $p(n)$ could be calculated by

$$p(n) = \sum_{k=0}^{N} a_k n^k, \tag{2.1}$$

that minimize the mean-squared approximation error for the group of input samples centered on $n = 0$

$$\xi_N = \sum_{n=-M}^{M} (p(n) - x[n])^2 = \sum_{n=-M}^{M} (\sum_{k=0}^{N} a_k n^k - x[n])^2. \tag{2.2}$$

Analysis is the same for any other group of $2M + 1$ input samples. M is referred as 'half width' of the approximation interval. When $N = 2$ and $M = 2$, the solid curve is shown on the left in Figure 2.1 and the polynomial $p(n)$ evaluated on a fine grid between -2 and +2, and the smoothed output value is obtained by evaluating $p(n)$ at the central point $n = 0$. The output at $n = 0$ is

$$y[0] = p(0) = a_0. \tag{2.3}$$

The output value is just equal to the 0th polynomial coefficient.

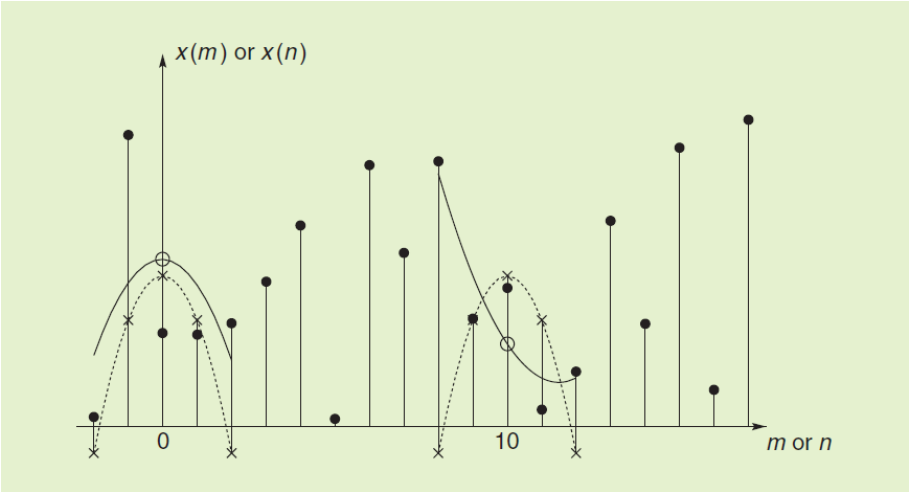

Figure 2.1: Illustration of least-squares smoothing [ZKM+18]

- • : as input samples,
- ◦ : as least-squares output sample,

$\times$: as effective impulse response samples (weighting constants).

Originally, SG filter is showed at each position [SG64]. The smoothed output value obtained by sampling the fitted polynomial is identical to a fixed linear combination of the local set of input samples: i.e., the set of $2M + 1$ input samples within the approximation interval are effectively combined by a fixed set of weighting coefficient that can be computed once for a given polynomial order N and approximation interval of length $2M + 1$. The output samples is calculated by discrete convolution from

$$y[n] = \sum_{n=-M}^{M} h[m]x[n-m] = \sum_{m=n-M}^{n+M} h[n-m]x[m]. \tag{2.4}$$

To find a single finite duration impulse response which is equivalent to least-square polynomial smoothing for all shifts of the $2M+1$ sample interval, coefficients of polynomial in equation (2.1) by different ξ_N in equation (2.2) with respect to each of the $N + 1$ unknown coefficient and setting the corresponding derivative equal to zero.

$$\frac{\alpha\xi_N}{\alpha a_i} = \sum_{n=-M}^{M} 2n^i(\sum_{k=0}^{N} a_k n^k - x[n]) = 0, \tag{2.5}$$

by interchanging the order of the summations, becomes the set of N+1 equations in $N + 1$ unknowns

$$\sum_{k=0}^{N}(\sum_{n=-m}^{M} n^{i+k})a_k = \sum_{n=-M}^{M} n^i x[n], i = 0, 1, ..., N. \tag{2.6}$$

Frame length and polynomial order which must be smaller than frame length are very important parameters in SG filter. In addition, parameters of weighting array and dimension to filter along are also have influence on the smoothing linear [WSS22].

2.1.3 Time-frequency analysis

Usually, brainwaves, vibration, and AE signals are oscillating motion, so most signals analysis approaches are used to determine the rate of that oscillation (frequency) [GP16]. Fourier transform (FT) is a common used mathematical technique that allows signal to be decomposed into a sum of sine waves of different frequencies, phases, and amplitudes. Frequency domain analysis helps to determine frequency band of the signals. Besides, it can often used for analyzing and designing nonlinear control systems. Although many advantages of domain analysis, disadvantages of frequency domain analysis are also very obvious: it struggles with nonlinear phenomena and lose the time information.

Time-frequency analysis (TFA) summarizes analysis techniques that quantify the temporal evolution of spectral properties of signals [Lee21]. These techniques provide powerful means to study the dynamics of rhythmic signals. Signals underlying changes in both time and frequency domain could be detected by time-frequency analysis, which can not be easily done within just the time domain or frequency domain. Time-frequency analysis methods applied in this study are Short-time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), and Hilbert-Huang transform (HHT).

Short-time Fourier Transform

Short-Time Fourier transform (STFT) is a sequence of Fourier Transform (FT) and provides the time-localized frequency information for situations in which frequency components of a signal vary over time [GC11]. The procedure of computing STFT is dividing a longer time signal into shorter segments of equal length and then compute the FT separately on all segments. According to [GC11], the STFT of a function $f(t)$ with respect to the window function $\phi(t)$ evaluated at the location (b, ξ) is defined as

$$G_\phi f(b,\xi) = \int_{-\infty}^{\infty} f(t)\overline{\phi_{b\xi}}\, dt, \tag{2.7}$$

with

$$\overline{\phi_{b\xi}(t)} = \phi(t-b)e^{j\xi t}, \tag{2.8}$$

using

$f(t)$	: as signal,
$\phi(t-b)$	: as window function,
b	: as desired location on time scale, and
ξ	: as desired location on frequency scale.

The window functions, denoted as sliding windows, are functions in which the amplitude tapers gradually and smoothly towards zero at the edges [LZM12]. The window function $\phi(t-b)$ keeps the desired portion of the signal by multiplication with the original signal $f(t)$. Outside of the desired interval, the window function is zero. By changing the parameter b, the window moves along the time axis. As a result, the local behavior of the signal can be analyzed. The most important parameters of the window related to center and width: in time domain, the center t^* and the root mean square radius Δ_ϕ while in frequency domain the center ω^* and the root mean square radius $\Delta_{\hat{\phi}}$. When its root mean square radius in time domain Δ_ϕ and in frequency-domain $\Delta_{\hat{\phi}}$ is finite, signal's time-frequency information is obtained. The time-frequency information of STFT is defined as the product of the time resolution and the frequency resolution [LT06]. The type of window affects the time-frequency resolutions of the STFT. Beside the window type, window length also affects the time and the frequency resolution of the STFT. According to [BYF+08], a narrow window results in a fine time resolution but a coarse frequency resolution as narrow windows have a short time duration but a wide bandwidth. A wide window results in a fine frequency resolution but a coarse time resolution as wide windows have a long time duration but a narrow frequency bandwidth. This phenomenon is called the window effect. Therefore, window type and parameters selection has a strong impact on signal's time-frequency resolution [WJS22].

Continuous Wavelet Transform

As stated by [GC11], the fixed time-frequency resolution of the STFT leads to serious constraints in many applications. For Instance, if a chirp signal with linear changing frequency over time should be transformed with the STFT, then depending on the chosen window, it could resolve either the lower frequencies but has a poor resolution on the higher frequencies or could resolve the higher frequencies but has a poor resolution on the lower

frequencies. A transformation that solves this problem is the Wavelet Transform (WT). The WT provides a window function whose radius increases in time (reduces in frequency) while resolving the low - frequency contents and decreases in time (increases in frequency) while resolving the high - frequency contents of a signal. This behavior is provided by a wavelet function $\Psi(t)$.

As stated by [GC11], The WT of a function $f(t)$ with the analyzing wavelet $\Psi(t)$ is defined as

$$W_\Psi f(b,a) := \int_{-\infty}^{\infty} f(t)\overline{\Psi_{b,a}(t)}\,dt, \tag{2.9}$$

where

$$\Psi_{b,a}(t) := \frac{1}{\sqrt{a}}\Psi(\frac{t-b}{a});\ a > 0, \tag{2.10}$$

with

- $f(t)$: signal,
- $\Psi(t)$: analyzing wavelet function,
- b : translation parameter on time scale, and
- a : dilation parameter on time and frequency scale.

Referring to [GC11], by reducing a, the support of $\Psi_{a,b}$ reduces in time and hence covers a larger frequency range and vice-versa. Therefore, $1/a$ is a measure of frequency. The parameter b indicates the location of the wavelet window along the time axis. Thus, by changing (b,a), $W_\Psi f$ can be computed on the entire time-frequency range. Hence the wavelet behaves like a window function.

In time domain, the center t^* and the root mean square radius Δ_Ψ are defined similar as STFT with $\phi(t)$ replaced by $\Psi(t)$ [GC11].

$$t^* = \frac{1}{||\Psi||^2}\int_{-\infty}^{\infty} t|\Psi(t)|^2\,dt, \tag{2.11}$$

$$\Delta_\Psi = \frac{1}{||\Psi||^2}\left[\int_{-\infty}^{\infty}(t-t^*)^2|\Psi(t)|^2,dt\right]^{1/2}. \tag{2.12}$$

Process of WT is similar to STFT, however, the window functions are different. Wavelet function increases its length and width when it shifts through the original signal while window function in STFT are stable. Window function difference between STFT and CWT are shown in Figure 2.2 Wavelet transform is usually divided into continuous wavelet transform (CWT) and discrete wavelet transform (DWT). As DWT is not applied in this study, so the emphasis is focus on CWT. Continuous wavelet transform is defined as

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t)\psi^*_{a,b}(t)dt. \tag{2.13}$$

In this formula, $x(t)$ refers to the impulse, a is scale parameter, b is shifting parameters, $\psi(t)$ refers to wavelet function. All variables and functions are determined in the realm of real number

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}). \tag{2.14}$$

The most noticeable characteristic change in wavelet from this formula is that the scale factor a was used to increase or decrease wavelet $\psi(t)$ and the translation factor b was used to move the wavelet [LZT95].

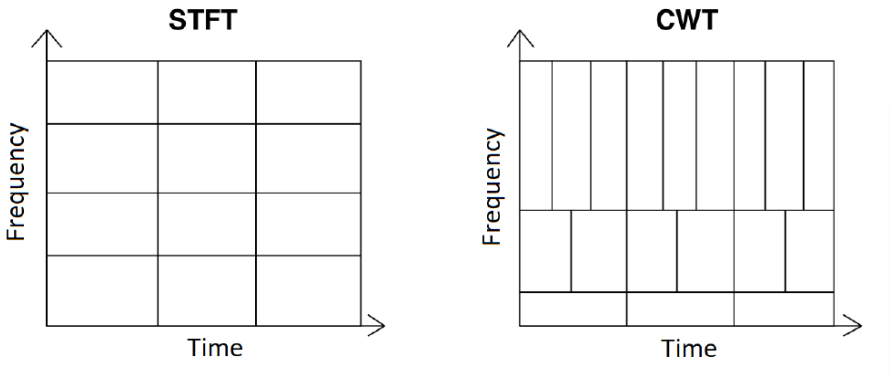

Figure 2.2: Window function of STFT and WT [YPZM19]

Hilbert-Huang transform

Hilbert-Huang transform is suitable on performing time-frequency analysis on nonstationary and nonlinear data [Bar11]. Two main steps are concluded in HHT: empirical mode decomposition and Hilbert spectral analysis. In the first step, the original signal is decomposed into a finite set functions which is known as intrinsic mode functions (IMFs) through an iterative process. The steps are:

1) Determine the local extreme of the signal;

2) Connect the maxima and minima with and interpolation function, generating an upper and a lower envelope about the signal;

3) Calculate the local mean as half the difference between the upper and lower envelopes. Subtract the local mean from the signal;

4) Iterate calculations on the residual.

Iterative process is repeated until the signal meets the definition of an IMF which is defined as signal with zero-mean, and its number of extreme and zero-crossing differ by at most one [WSS22].

Second step of HHI is transforming IMFs from time domain to time-frequency domain by Hilbert transform. According to [Joh99], the following equation is known for Hilbert transform

$$H[g(t)] = g(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{g(\tau)}{t-\tau} d_\tau = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{g(t-\tau)}{\tau} d_\tau, \tag{2.15}$$

with

$H[g(t)]$: Hilbert transform,

$g(t)$: original signal.

Iterative process is repeated until the signal meets the definition of an IMF which is defined as signal with zero-mean, and its number of extreme and zero-crossing differ by at most one.

2.1.4 Data normalization

Data normalization is used to flatten value differences or predictors inside same classes. One advantage of data normalization is that data get rid of anomalies. Besides, data

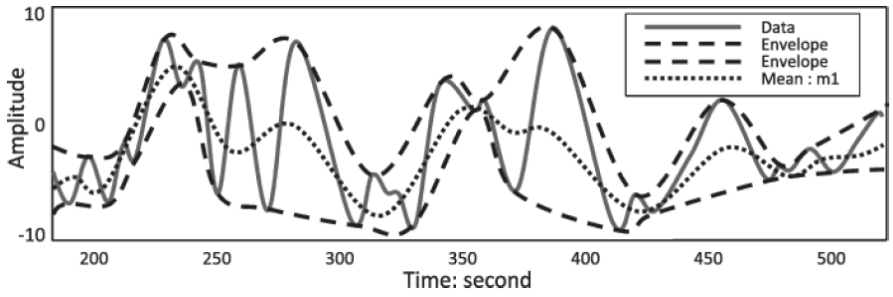

Figure 2.3: Iterative sifting process [LHCT18]

normalization reduces database space requirements. In the proposed approach, Z-Score and Min-Max techniques are applied for data normalization [WJS22].

Z-Score normalization

Z-Score normalization uses the mean and the standard deviation of the given data. As result the data has a mean of 0 and a standard deviation of 1. As stated by [JNR05], The normalized scores (S'_k) are given by

$$S'_k := \frac{(S_k - \mu)}{\sigma}, \tag{2.16}$$

where S_k is the raw data, μ is the arithmetic mean, and σ is the standard deviation of the given data. According to [JNR05], if the input scores are not Gaussian distributed, Z-Score normalization does not retain the input distribution at the output. This is due to the fact that mean and standard deviation are the optimal location and scale parameters only for a Gaussian distribution.

Min-Max Normalization

The Min-Max Normalization is a linear transformation, that rescales the data to defined boundaries [SS20]. Common boundaries for rescaling are [0-1]. As stated by [SS20], the Min-Max Normalization for a value set A with the rescale boundaries $[B_{min}, B_{max}]$ is defined as

$$A' := \frac{A - A_{min}}{A_{max} - A_{min}} * (B_{max} - B_{min}) + B_{min}, \tag{2.17}$$

with

A : as original value set,

A_{min} : as minimum value of A,

A_{max} : as maximum value of A,

A' : as rescaled value set of A,

B_{min} : as minimum value of A', and

B_{max} : as maximum value of A'.

2.2 Machine learning

Both supervised and unsupervised learning approaches are employed in this study. The difference between supervised learning and unsupervised learning approaches is: supervised learning used labeled data to help predict outcomes while the unsupervised learning employed unlabeled data. Supervised learning models are defined by labeled datasets. These datasets designed to train or 'supervise' algorithms into classifying data or predicting outcomes accurately [Mah20]. Supervised learning can be categorized into two types in the process of data mining: classification and regression. Unsupervised learning algorithms are used to analyze and cluster unlabeled datasets and discover hidden patterns in data without human intervention [SS18]. Unsupervised learning models are used for the tasks of clustering, association, and data dimension reduction. Transfer learning (TL) is machine learning method where a model developed for a task is reused as the starting point for a model on a second task [ADC21]. Supervised learning, unsupervised learning approaches, and transfer learning relevant to this study are introduced in this section.

2.2.1 Supervised learning

Supervised learning algorithms are the most common type of machine learning algorithms. It uses a known dataset (called the training dataset) to train an model and to make predictions [SK18]. The training dataset includes labeled input data that pair with desired outputs or response values. Supervised learning algorithm seeks to create a model by discovering relationships between the features and output data, then makes predictions of the response values for a new dataset [Lan19].

Convolution Neural Network

Convolution neural network is the most famous and commonly used algorithm in field of DL. It has the ability to automatically learn a large number of filters in parallel specific to a training dataset under the constraints of a specific predictive modeling problem [AZH$^+$21]. As stated in [AZH$^+$21], the CNN algorithm extracts features from input data and classifies them into predefined classes based on trained and extracted features. A CNN model generally consists of three parts. The first part is the image input with the image input layer. The second part is related to feature extraction, consisting of numerous convolution layers with preceding sub sampling (pooling) layer and subsequent non-linearity activation functions [APG15]. In addition, batch normalization and dropout layers are used to prevent overfitting and other issues. The third part classifies the input image based on trained and extracted features. It consists at least one fully connected (FC) layer followed by a loss function layer.

The convolution layer is the most important layer of the CNN architecture. It consists of convolution filters named kernels [PSJL17]. Kernels which are used to extract the features from input data are low in size and often have the same dimensions as the input. The input data are convolved with these kernels and outputs denoted as feature maps [BS16]. Feature maps represent the automatically extracted features of the input data. To get precise and detailed features, numerous convolution layers are used in hierarchical order. The basic principle of the construct convolution layer is shown in Figure 2. When input elements i_{mn} which belong to part of the input data matrix $[I]$ $(i_{11}, i_{12}, ..., i_{mn})$ dot

product with kernel $[K]$ $(k_{11}, k_{12}, ..., k_{mn})$, output matrix $[O]$ $(o_{11}, o_{12}, ..., o_{mn})$, which has the same size as kernel is gotten.

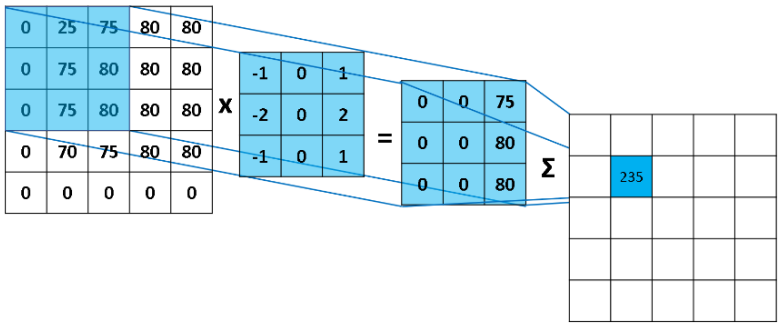

Figure 2.4: Convolution process [Wu17]

The convolution process is the following form:

$$O = I \cdot K = \begin{bmatrix} i_{11} & i_{12} & i_{13} \\ i_{21} & i_{22} & i_{23} \\ i_{31} & i_{32} & i_{33} \end{bmatrix} \cdot \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} = \begin{bmatrix} 0 & 25 & 75 \\ 0 & 75 & 80 \\ 0 & 75 & 80 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 75 \\ 0 & 0 & 80 \\ 0 & 0 & 80 \end{bmatrix},$$

which is shown in the Figure 2.4. The number is then transformed to the next layer when all elements of $[O]$ $(o_{11}, o_{12}, ..., o_{mn})$ in output matrix are summed [WJS22].

Batch normalization layer make neural networks faster and more stable. It performs the standardizing and normalizing operations on the input of a layer coming from a previous layer [HZZ+19]. Advantages of batch normalization are as follows:

i) Prevention from vanishing gradient problem,

ii) Effective control of poor initialized weights,

iii) Significantly reduction of the time required for network convergence,

iv) Decreasing of training dependency across hyerparameters, and

v) Reduction of overfitting through regularization.

Activation function layer is non-linear activation layer which are employed after each convolutional layer [AZH+21]. These layers map their input to their output with a non-linear function. Goal of this layer is to give CNN the ability to learn complicated structures and to differentiate which is an extremely significant feature. The most common used activation functions are sigmoid, tanh, and ReLU functions.

Pooling layer is used to sub-sample the feature maps produced by convolutional layer [AFAS+20]. Large-size feature map is shrinked to smaller feature map by pooling operation. Majority of the dominant information are maintained while overall memory usage are reduced. Most common used pooling techniques are: max pooling, average pooling, global average pooling.

Dropout is a common technique for generalization as stated by [AFAS+20]. In each epoch, randomly chosen neurons are dropped. Dropped neurons are not considered for forward- or backward- propagation. With dropping randomly chosen neurons, the feature selection power is equally distributed over all neurons and the model is forced to learn different independent features. Dropout layer also prevents model from overfitting.

Fully connected layer located at the end of CNN consists of neurons that are connected to all neurons to all neurons of the previous layer [AZH+21]. The last FC layer has as

many neurons as the class number that need to be classified. To fit the N-dimensional feature maps from the feature extraction part of the CNN to the FC layer, the feature maps have to be flattened. Flattening layer which maps all values into one vector and by global pooling layer are employed to flatten feature maps.

Cross entropy or softmax loss function is usually applied for multi-class classification in the output layer [AZH+21]. It outputs the probability $p \in \{0, 1\}$ for each class $i \in [1, N]$. The softmax output class probability is defined as

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}, \tag{2.18}$$

with

e^{a_i} : Non-normalized output from the preceding layer,

N : Number of neurons in the output layer (number of classes).

The cross entropy loss function is defined as

$$H(p, y) = -\sum_i y_i \log(p_i), \tag{2.19}$$

with

p_i : Predicted output,

y_i : Desired output.

Support vector machine

Support vector machine is one of the most popular supervised learning algorithms primarily used for classification problems. Goal of SVM is to generate the best line or decision boundary that can segregate n-dimension space into classes so that new data can be easily sorted in the correct category [TW04]. The best decision boundary is named as hyperplane. Support vector machine chooses the extreme points/vectors which are called support vectors to help in generating the hyperplane.

Support vector machine can be categorized into two types: linear and non-linear [GDS19]. Mapping from the input space into the feature space is explained as well as the 'kernel trick' in SVM. A kernel function can be interpreted as a kind of similarity measure between input objects [GD05]. Various kernels can be used such as: linear, polynomial, and Gaussian kernels. In the proposed approach, linear kernel, different order polynomial kernel, and varied Gaussian kernels are tried.

The linear kernel function is

$$k(x_i, x_j) = 1 + x_i^{'} x_j. \tag{2.20}$$

The polynomial kernel function is

$$k(x_i, x_j) = (1 + x_i^{'} x_j)^p. \tag{2.21}$$

The Gaussian kernel function is

$$k(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{2\delta^2}). \tag{2.22}$$

The denotation of parameters in the above three equations are

x_i, x_j : features or data points,

p : polynomial order,

δ : kernel scale.

According to the δ calculation formula, Gaussian SVM can be divided into fine Gaussian, medium Gaussian and coarse Gaussian.

2.2.2 Unsupervised learning

Unsupervised learning algorithms perform more complex processing tasks compared with supervised learning. It finds all kind of unknown patterns in data and features which can be useful for categorization.In addition, all input data can be analyzed and labeled in the presence of learners by unsupervised learning. Unsupervised learning problems further grouped into clustering and association problems.

Autoencoder

Autoencoder is a type of neural network used to learn data encodings in an unsupervised manner [Lee21]. The aim of autoencoder is to learn a lower-dimension representation for a higher-dimensional data, typically for dimensionality reduction by training the network to capture the most important parts of the input image [VBR20]. Architecture of autoencoder consists three parts as shown in Figure 2.5:

i) Encoder: compress input data into an encoded representation which is typically several orders of magnitude smaller than the input data.

ii) Bottleneck: a module that contains the compressed knowledge representations and is the most important part of the network.

iii) Decoder: decompress the knowledge representations and reconstructs the data back from its encoded form. The output is then compared with a ground truth. According to [BAR20], advantages of autoencoder are as follows:

i) Encoding part of the architecture is helpful to reduce the complexity of input data. By reducing number of input values,the model is less likely to be overfitting by tiny details.

ii) When data are not labeled, autoencoder find the new classes. This is especially useful for unsupervised learning or clustering.

iii) Autoencoder is very powerful for detecting anomalies.

K-means

K-means clustering is one of the typical and popular unsupervised machine learning algorithms. This algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group [VS20]. It tries to groups objects which are close together into the same cluster while the objects that are further apart into different clusters [ECS11]. Data are signed to a cluster in this way: the sum of the squared distance between data points and the cluster's centroid is at the minimum. As shown in Figure 2.6, the process of K-means can be divided into following steps:

i) Initial cluster centers which known as centroids are placed randomly.

ii) Compute the sum of the squared distance between data points and all centroids.

iii) Assign each data point to the closest cluster.

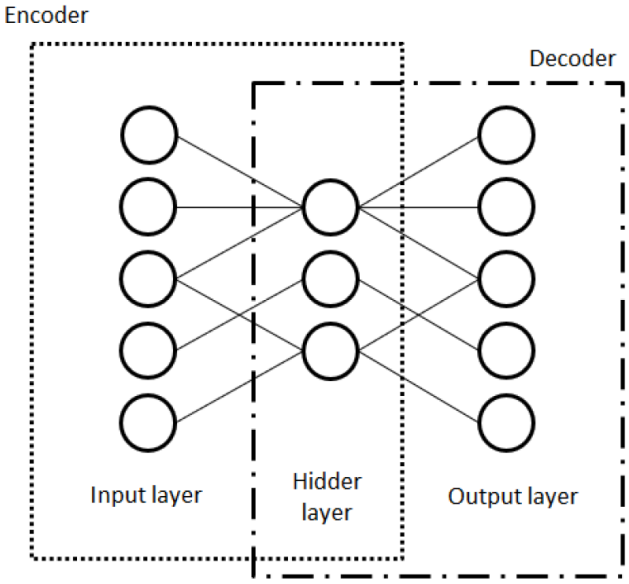

Figure 2.5: Architecture of autoencoder [SK19]

iv) Compute the centroids for the clusters by taking the average of all data points that belong to each cluster.

v) Repeat the previous steps until cluster assignment is constant or maximum number of iterations is reached. The approach K-means follows to solve the problem is called

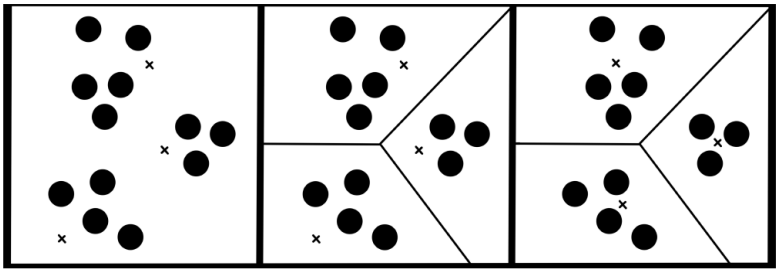

Figure 2.6: Process of K-means [ECS11]

Expectation-Maximization [BFR+98]. E-step assigns data points to the closest cluster while M-step compute the centroid of each cluster. The objective function is

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \|x^i - \mu_k\|^2, \tag{2.23}$$

with

x^i : as input data point,

μ_k : as the centroid.

Data point x^i belong to the cluster K, when w_{ik}=1. Otherwise, w_{ik}=0. Expectation-Maximization algorithm addresses problem in two steps. Firstly, w_{ik} is minimized and μk

is fixed and then μk is minimized and w_{ik} is fixed. These steps are done in E-step and M-step separately. In E-step, data point x^i are assigned to the closest cluster judged by its sum of squared distance from cluster's centroid. Equation of E-step is as follows

$$\begin{aligned} \frac{\alpha J}{\alpha w_{ik}} &= \sum_{i=1}^{m}\sum_{k=1}^{K} \|x^i - \mu_k\|^2 \\ \Rightarrow w_{ik} &= \begin{cases} 1, & \text{if} \quad k = argmin_j \|x^i - \mu_k\|^2 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.24}$$

In M-step, centroid of each cluster is recomputed to reflect the new assignments. Equation of M-step is

$$\begin{aligned} \frac{\alpha J}{\alpha \mu k} &= 2\sum_{i=1}^{m} w_{ik}(x^i - \mu_k) = 0 \\ \Rightarrow w_{ik} &= \frac{\sum_{i=1}^{m} w_{ik}(x^i}{\sum_{i=1}^{m} w_{ik}}. \end{aligned} \tag{2.25}$$

After iterative calculation, stop conditions for K-means could be as follows [GT14]:

i) Data points assigned to specific cluster remain the same;
ii) Centroids remain the same;
iii) The distance of data points from their centroid is minimum;
iv) Fixed number of iterations have reached.

Gaussian mixture models

Gaussian mixture models are a probabilistic model that assumes all the data points are generated from a mix of Gaussian distribution with unknown parameters. As stated in [LZMZ18], GMMs can be applied to find clusters in datasets where the clusters may not be clearly defined. In addition, GMMs can also be used to estimate the probability of new data point belongs to each cluster. Furthermore, GMMs are relatively robust to outliers, this means that they can output accurate results even if some data points unfit for nearly any of the clusters.

According to [Rey09], a GMM is a weighted sum of M component Gaussian densities as given by the equation,

$$p(X|\lambda) = \sum_{i=1}^{M} \omega_i g(X|\mu, \sigma_i). \tag{2.26}$$

Each component density is a D-variate Gaussian function of the form,

$$g(X|\mu_i, \sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}}|\sigma_i|^{\frac{1}{2}}} exp\{-\frac{1}{2}(X - \mu_i)' \sigma_i^{-1}(X - \mu_i)\}. \tag{2.27}$$

The parameters signification in the above two equations are

X	: as D-dimensional continuous-valued data vector,
ω_i	: as mixture weights,
μ_i	: as mean vector,

σ_i : as covariance matrix,

$g(X|\mu, \sigma_i)$: as component Gaussian densities.

The mixture weights satisfy the constraint of

$$\sum_{i=1}^{M} \omega_i = 1. \tag{2.28}$$

The complete GMM is parameterized by the mean vectors, covariance matrices, and mixture weights from all component densities. These parameters are collectively represented by

$$\lambda = \{\omega_i, \mu_i, \sigma_i\} \quad i = 1, 2, ..., M. \tag{2.29}$$

In [NRBZ15], GMMs are also assumed that there are a certain number of Gaussian distribution and each of these distribution represent a cluster. therefore, a GMM tends to group data points belonging to a single distribution together. Given three Gaussian distributions (GD) - GD1, GD2, and GD3. These have a certain mean ($\mu 1, \mu 2, \mu 3$) and variance ($\sigma 1, \sigma 2, \sigma 3$) value respectively. For a given set of data point, GMM assigns Gaussian distribution to each cluster as shown in Figure 2.7. The bell curve represents a GD of a cluster and the dots are data points. Unlike K-means, GMMs can perform both hard and soft clustering. In hard clustering, each data point is assigned to only one cluster while soft clustering data point can be assigned to multiple clusters based on the probabilities.

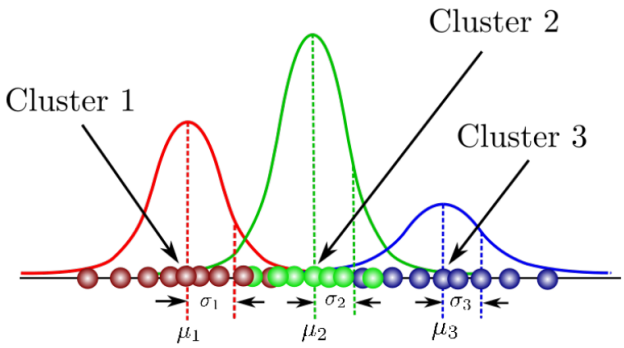

Figure 2.7: Gaussian distribution [NRBZ15]

2.2.3 Transfer learning

Transfer Learning is a machine learning technique where a model trained on source domain is repurposed on target domain. According to [WKW16], some notations and definitions used in TL are introduced. The definition of 'domain' and 'task' is defined in the following. According to [SKS16], a domain D consists of two components: a feature space Ω and a marginal probability distribution $P(\Omega)$, where $\Omega = \{x_1, x_2, \ldots, x_n\} \in \Omega$. Given a specific domain, $D = (\Omega, P(\Omega))$, a task consists of two components: a label space y and an objective predictive function $f(.)$ (denoted by $T = (y, f(.))$), which is not observed but can be learned from training data, which consists of pairs $\{x_i, y_i\}$,where $x_i \in \Omega$ and $y_i \in y$. the function $f(.)$ can be used to predict the corresponding label $f(x)$ of a new instance x. Given a

source domain D_S and learning task T_S, a target domain D_T and learning task T_T, transfer learning aims to improve the learning of the target predictive function $f_T(.)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$ [SKS16]. The process of TL is shown in Figure 2.8. Based on different situations between source and target domains

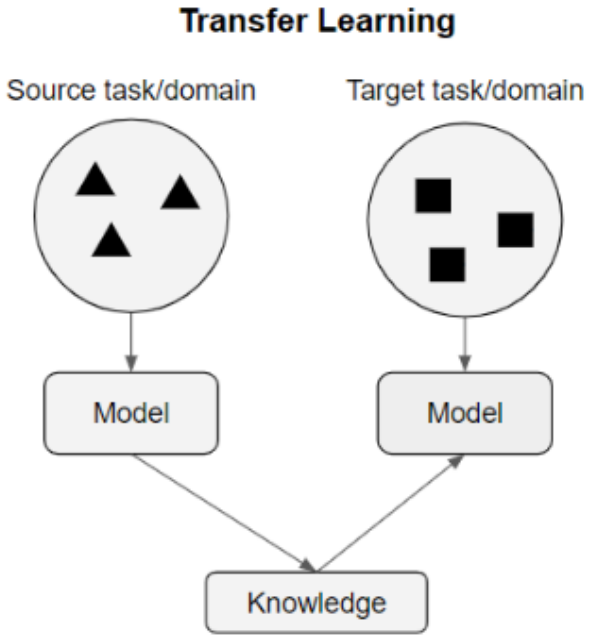

Figure 2.8: Transfer learning [Rud19]

and tasks, TL can be categorized in three subsettings: inductive TL, transductive TL, and unsupervised TL [WKW16]. The inductive Transfer Learning setting, the target task is different from the source task, when the source and target domains are the same. In the transductive TL setting, the source and target tasks are the same, while the source and target domains are different [WJDS22]. In the unsupervised TL setting, similar to inductive TL setting, the target task is different from but related to the source task [ZHWL14]. However, the unsupervised TL focuses on solving unsupervised learning tasks in the target domain, such as clustering, dimensionality reduction, and density estimation. Based on 'what to transfer', approaches to TL in the above three different settings can be summarized into four cases: instance-transfer, feature-representation-transfer, parameter-transfer, and relational-knowledge-transfer.

2.3 Parameter and hyperparameter optimization algorithms

A large amount of different hyperparameters and corresponding values are need to optimized during machine learning approaches. To automatically train a model with various initial conditions, Experiment Manager is applied in this study. The Experiment Manager is used in this thesis to find optimal options for segmentation, transformation, and hyperparameters in machine learning [Joc21]. To find these optimal options, both exhaustive sweep and Bayesian optimization are applied in this study.

2.3.1 Exhaustive sweep

The exhaustive sweep algorithm is used to sweep to all possible combinations of parameter values that need to be settled. For example the initial learning rate need to be optimized in CNN. The values for that parameter the algorithm should sweep through could be $[0.0001, 0.001, 0.01]$. Besides the initial learning rate, variety of hyperparameters need to optimize. The more differently parameter and corresponding values are set, the more combinations of all parameter values are possible. In this context and considering the computational speed of training, the algorithm should only be applied to a greater amount of parameters with few values or a small amount of parameters with larger value sets. Otherwise the number of experiments gets to large to be computed in a reasonable amount of time [Joc21]. Although this algorithm is a good option to get a rough overview of good parameter ranges, it is not suitable to determine the optimal parameter values precisely. To precisely determine the optimal parameter values from a set of roughly determined good parameter ranges, the Bayesian Optimization algorithm is needed.

2.3.2 Bayesian optimization

The Bayesian optimization algorithm minimizes a scalar objective function [ISN+18]. This function is represented in the context of parameter optimization by one of the evaluation metrics of the trained models. This could be for example the validation accuracy or the training loss. The Bayesian optimization algorithm minimizes the distance of the evaluation metric from its optimal value by changing the initial parameter values in a given range. The probabilistic model for the scalar objective function is a Gaussian process prior with added Gaussian noise.

Different from the Exhaustive Sweep algorithm, no concrete parameter values but a value range for each parameter is set in Bayesian optimization. The Bayesian optimization algorithm has no trials on each combination of the given parameter values, but it finds the values that optimize the evaluation metric in a given time or number of experiments. For large value ranges and a larger amount of parameters this algorithm needs lot amount of experiments to optimize the evaluation metrics. Therefore this algorithm should be used to fine tune roughly determined good parameter ranges of a small number of parameters [Joc21].

2.4 Evaluation metrics

Models performance should be evaluated after they are trained. Evaluation metrics ae used to measure the quality of the statistical or machine learning models. A large number of evaluation metrics are available to test a model. According to ML tasks, evaluation metrics could be divided into metrics for supervised learning approaches and metrics for unsupervised learning approaches.

2.4.1 Metrics for supervised learning

Following [AZH+21], to achieve an optimized classifier, diverse metrics play a crucial role for evaluation of ML approaches. Evaluation metrics are utilized through the two main stages of the generation of a usual ML classifier: training and test. The optimized solution of the training stage is discriminated and selected using the evaluation metric. Therefore,

evaluation metrics are utilized to measure the efficiency of the generated class-related model. For test, the selected classifier gets evaluated considering the evaluation metric. Then results for test stage and the efficiency on training stage are compared to show the classifier performance on untrained data.

To evaluate a trained model performance, the classified data have to be categorized into true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Numbers of TP and TN are the samples number that are correctly classified [AZH+21]. Numbers of FP and FN are the samples number that have been misclassified [AZH+21]. The most common metrics for classification are accuracy, recall, precision, and F-score. Accuracy calculates the ratio of correct predicted cases to the overall number of evaluated samples. Recall or Sensitivity calculates the ratio of TP over the total number of TP and FN. Precision calculates the ratio of TP over the total number of TP and FP. F-Score] calculates the harmonic average between recall and precision [WS20]. Every metric has its specific pros and cons: accuracy assesses the overall effectiveness of algorithms, precision assesses the predictive power of algorithms, sensitivity and specificity access the effectiveness of the algorithm on a single class; F-score benefits algorithms with higher sensitivity and challenges algorithms with higher specificity.

To evaluate the proposed approach comprehensively, models are tested in two schemes: individual scheme and all subjects scheme. In individual model, both training and test data come from the same subject. In all subject scheme, training data come from all subjects and test data come from part of each subject [Sur22]. To measure the distribution of models trained by each subject, standard deviations (SD) is applied as

$$SD = \sqrt{\frac{|x - \hat{x}|^2}{n}}. \tag{2.30}$$

In statistics, SD is the degree of dispersion or the scatter of the data points relative to its mean. It illustrates the values are spread across the data sample and it is the measure of the variation of the data points from the mean [DL05].

2.4.2 Metrics for unsupervised learning

Two types of metrics (internal and external metrics) can be employed for unsupervised learning models (especially clustering models). Internal metrics define the quality of a clustering without external labels by using the idea of cohesion and separation. External metrics are understood as an equivalent the evaluation metrics of supervised algorithms. Data labels must be available for external metrics. As the AE data are labeled in this study, external metrics are applied [Lee21]. To evaluate the approach comprehensively, five metrics – purity, rand index (RI), adjusted rand index (ARI), normalized mutual information (MNI), and F-measure – are employed.

Among external metrics, purity is a simple and transparent evaluation measure. It is the percent of the total number of objects that were classified correctly [MGB+11]. Rand index computes a similarity measure between two clustering by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clustering [BRK+18]. The drawback of RI is yielding a high value for pairs of random partitions of a given set of examples. Adjustment rand index is employed to the calculation by taking into consideration grouping by chance to counter RI drawback. Normalized mutual information measure the similarities between two labels of

the same data and balance the cluster quality and cluster numbers [SG02]. The advantage of NMI is that it can be applied to compare different cluster models that have different number of clusters. F-measure is a balanced mean between precision and recall and it benefits algorithms with higher sensitivity and challenges algorithms with higher specificity [NBP09].

3 Application fields and related datasets

Three kinds typical signals: EEG, vibration, and AE signals which belong to various application fields are analyzed in this study. Electroencephalogram signals belong to biological field. They are one kind of brainwaves with low frequency (from 0 to hundred Hz). Vibration is a mechanical event in which oscillations occur about an equilibrium point. The time series that carries the information of those oscillations is called vibration signal [RS16]. Vibration signals have a wide range of applications and can be used in almost every industry, such as paper, coal and construction industries. Vibrations can be described both in intensity by amplitude and in periodicity by frequency. Frequency bands of vibration signals are very wide (from dozen to thousand Hz). Acoustic emission or stress wave emission is the phenomenon of transient elastic wave generation due to a rapid release of strain energy caused by a structural alteration in a solid material [Gre80].It is a very efficient and effective technology used for fracture behavior and fatigue detection in metals, fiberglass, wood, composites, ceramics, concrete and plastics [GLB15]. Depending on the source mechanism, acoustic emission signals may occur with frequencies ranging from several hertz up to tens of megahertz. While these three kind of signals characteristics are varied and they have wide application in different fields, they are employed in this study. Signals and corresponding datasets applied will be introduced briefly in this chapter.

The contents, figures, and tables presented in this chapter are modified after previous publications [WS20], [WJDS22], and [WDS21]. Part of the contents, figures, and tables are prepared for publications [WSS22], [WLS22] and [WJS22].

3.1 Inner Speech (IS) dataset

Inner speech (IS) (silent-, covert-, speech, verbal thought) are some of the terms used to refer to the silent production of words in one's mind, or the activity of talking to oneself silence [ADF15]. It is defined as the silent expression of conscious thought to oneself in a coherent linguistic form [PBRL$^+$14]. Though much difficult in studying inner speech, the role it plays in different cognitive abilities, including memory and executive functioning is well established. It is linked to the development of language abilities and the advanced mental abilities to which language is linked [ESS05]. Another skill that appears linked to inner speech is silent reading. In addition, inner speech helps certain brain disorders resulting from brain stem infarcts, traumatic brain injury, cerebral palsy, stroke, and amyotrophic lateral sclerosis, limit verbal communication despite the patient being fully aware [JG14]. People that cannot communicate due to neurological disorder would benefit from systems that can infer internal speech directly from brain signals. People with speech deficits would benefit from a communication system that can directly infer inner speech from brain signals – allowing them to interact more naturally with the world [MIM$^+$18]. Therefore, inner speech recognition has been proposed as an alternative communication

paradigm for brain-computer interface (BCI) [PR21].

Brain-computer interface is a collaboration between brain and device that enables signals from the brain to direct some external activity. The interface enables a direct communication pathway between the brain and the object to be controlled [CGGX02]. In addition to its application in medical and health field, BCI technology has potential applications in military, education, and recreation fields. According to [AAM15], BCI system consists of four basic components: signal acquisition, signals preprocessing, feature extraction, and classification. Discriminative characteristics of the improved signals are extracted in features extraction stage. Lastly, classifiers distinguish features and allow finally the guidance of device commands [WSS22].

Signal acquisition is measuring brain oscillating electrical voltages which is also named brain waves [AGM16]. Brain waves reflects the voluntary neural actions generated by user's current activity. Methods capturing brain waves can be divided into invasive and non-invasive. Invasive recording methods implant electrodes under the scalp and measure the neural activity of the brain either intracortically from within the motor cortex or on the cortical surface (electrocorticography (ECoG)) [AAM15]. The most relevant advantage of invasive recording is providing high temporal and spatial resolution, increasing the quality of the obtained signal, and a good signal to noise ration [AAM15]. Main downside is the utilization of invasive surgery, and the potential for scar tissue to form around the site, which may lead to potential side effects such as seizures [MMM$^+$16]. The invasive systems are mostly used in BCI systems experiment that use monkeys according to [BAB14]. Instead of the surgical procedure and permanent device attachment, non-invasive recording methods record brain activity from electrodes placed on the skin and scalp. In general, non-invasive are considered the safest and low-cost type of devices. Functional magnetic resonance imaging (fMRI), functional near infrared spectroscopy (fNIRS), magnetoencephalography (MEG), electroencephalogram (EEG), and stereotactic electroencephalography (sEEG) belong to the non-invasive methods [HBS$^+$17]. Among these non-invasive recording methods, EEG is one of the most common used methods in BCI system. It is a physiological method of choice to record the electrical activity generated by the brain via electrodes placed on the scalp surface [T$^+$02]. These electrodes can only capture 'weaker' human brain signals due to the obstruction of the skull.

A large number of research focuses on EEG-based BCI system [ZTS$^+$18, LCL$^+$07, ANHAOAW17, LBC$^+$18, HKCI13]. Common BCI paradigms, the signal processing, and feature extraction methods are introduced in [ZTS$^+$18]. In [LCL$^+$07, LBC$^+$18], commonly employed algorithms used to design BCI systems based on EEG are presented and relevant critical properties are described. The state-of-art as well as trends in EEG-based emotion recognition system research are summarized by AI-Nafjan et al. [ANHAOAW17] focusing on emotion detection, recognition, and classification. Brain-computer interface paradigms, EEG feature types, classification algorithms, and target applications from 2007 to 2011 are revealed in [HKCI13]. In other words, data processing and distinguishing approaches are summarized and reviewed in these contributions. Channel selection, time window setting, and artifacts removal are employed in data preprocessing component. For feature extraction component, motor imagery (MI)-based EEG, steady state visual evoked potentials (SSVEPs), steady-state somatosensory evoked potentials (SSSEPs), and P300 are applied. Finally, features can be distinguished by linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbors (k-NN), and artificial neural networks (ANNs) [WSS22]. The current review evaluates EEG-based BCI paradigms regarding their

advantages and disadvantages from a variety of perspectives. For each paradigm, various EEG decoding algorithms and classification methods are evaluated and their application [ABS+19].

Owing to lack of publicly available electroencephalography datasets restricts the development of new techniques for inner speech recognition, Nieto et al. [NPR+22] conducted an experiment to provide an open-access multiclass electroencephalography database of inner speech commands to the scientific community. In the experiment, ten healthy right-handed subjects – four females and six males with mean age ± std = 34 ± 10 years – without any hearing or speech loss, participated in the experiment. Electroencephalography headcap with 128 electrode were placed on each subject before the experiment. Besides headcap electrode, two computers and one computer screen graphic user interface (GUI) are included in the test as shown in Figure 3.1.

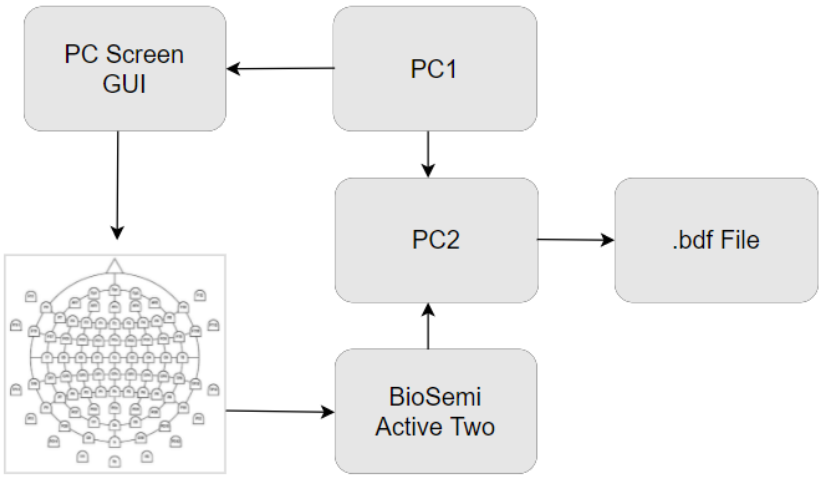

Figure 3.1: Test rig [WSS22]

Each subject participated in one single recording day comprising three consecutive sessions. Within each session, five stimulation runs were presented. Four trial classes (words) – 'Arriba (up)', 'Abajo (down)', 'Derecha (right)', and 'Izquierda (left)' – are selected and presented in the screen. The trial's class (word) is selected randomly. The occurrence frequency of each word is the same, 559 times for each word. In total, 2236 trials are held for inner speech. The detailed trial number of each subject is shown in Table 3.1.

According to [NPR+22], the procedure of each trial is as follows. Each trial begin at time $t = 0$ s with a concentration interval of 0.5 s. A white circle appears in the middle of the screen and participant should fix his/her gaze on it. Participants should not blink until the white circle disappear. At time $t = 0.5$ s the sign interval started. A white triangle pointing to different direction is presented. The pointing direction of the sign corresponds to each class. At $t = 1$ s, the triangle disappears from the screen and only the white circle remains. At the same time, the action interval started. After the visual cues disappeared, participants are instructed to imagine his/her own voice as if he/she was giving a direct order to the computer, repeating the corresponding word. After 2.5 s of action interval, i.e. at $t = 3.5$ s, the white circle turned blue and relax interval began. At $t = 4.5$ s the blue

Table 3.1: Inner speech trials [Sur22]

Subjects	Up	Down	Right	Left
sub-01	50	50	50	50
sub-02	60	60	60	60
sub-03	45	45	45	45
sub-04	60	60	60	60
sub-05	60	60	60	60
sub-06	54	54	54	54
sub-07	60	60	60	60
sub-08	50	50	50	50
sub-09	60	60	60	60
sub-10	60	60	60	60
Sub total	559	559	559	559
Total		2236		

circle vanished and one trial ended. The workflow of each trial is shown in Figure 3.2.

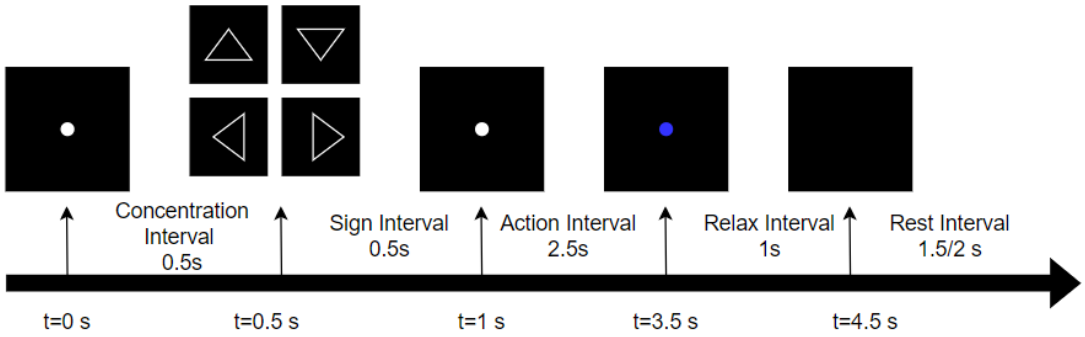

Figure 3.2: Trial workflow [WSS22]

3.2 Case Western Reserve University (CWRU) bearing dataset

Within the last decades, rotating machinery equipment plays an irreplaceable role in modern industry [ZLW+20]. As one of the most common components of rotary machinery, bearing is a mechanical component used to reduce friction among other moving parts. Once a bearing fails (or components in it), other adjacent components and machines are effected up to failure. Several surveys regarding the likelihood of induction machine failure conducted by the IEEE Industry Application Society (IEEE-IAS) and the Japan Electrical Manufactures' Association (JEMA) reveal that bearing fault is the most common fault type and is responding for 30 to 40 % of all machine failures [ZZWH20]. Therefore, condition monitoring and fault diagnosis of bearings is of increasing interest [BMDA16]. Several benchmark datasets are developed to evaluate development in bearings health state (diagnosis) and remaining useful lifetime (prognosis). Among these datasets, Case Western Reserve University (CWRU) dataset is one of the most cited ones used to validate the performance of different approaches on bearing diagnosis [WS20].

Case Western Reserve University bearing dataset is generated by Prof. Kenneth Loparo's research group and provides access to ball bearing test results for fault-free and faulty bearings [WS20]. As shown in Figure 1, the test rig consists of a 2 Hp motor on the left side, a torque transducer in the center, a dynamometer on the right side, and control electronics (not shown) [WS20]. The test bearings support the motor shaft and vibration data were recorded for different motor loads (labeled as 0, 1, 2, 3, corresponding motor speed is 1797, 1772, 1750, 1720 rpm). A single point fault is introduced to the test bearings at drive end (DE) and fan end (FE) using electro-discharge machining with fault diameters of 007, 014, 021, and 028 mils (1 mil = 0.001 inch) [WS20]. Faults are introduced separately at the inner raceway (IR), ball (B), and outer raceway (OR). Outer raceway faults are located at 3 o'clock (directly in the load zone), 6 o'clock (orthogonal to the load zone), and at 12 o'clock [WS20]. Vibration data are collected using accelerometers. Accelerometers are placed at the 12 o'clock position at both the drive end and fan end of the motor housing. During some experiments, an accelerometer is attached to the motor supporting base plate (BA) as well. Digital data are collected at samples rate of 12,000 samples/second. Data are also collected at 48,000 samples per second for faulty bearings at drive end. Four sub-databases are included in CWRU bearing dataset: normal baseline

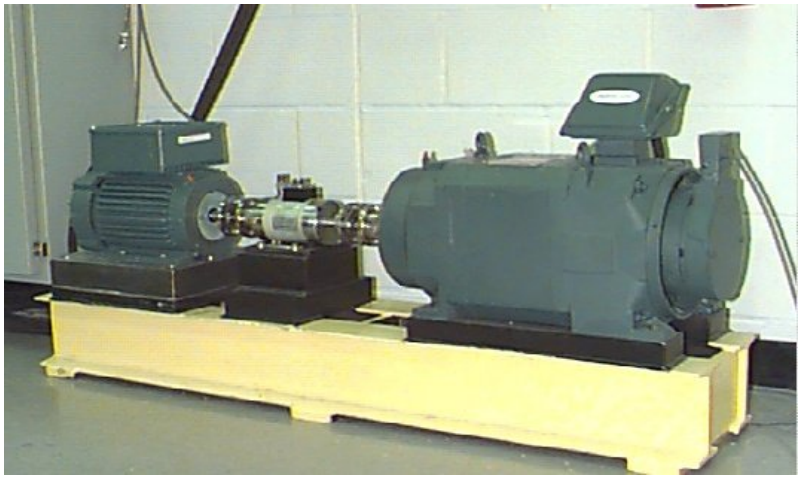

Figure 3.3: Test rig of CWRU bearing dataset [WS20]

(fault-free) data, 12k and 48k drive end bearing fault data, and fan-end baring fault data. For drive end bearing states, same bearing states are recorded in both 12k and 48k drive end bearing sub-databases, the differences between them are related to the sampling rate. Depending on the located position of faulty bearing, fault size, and fault location in outer raceway, 28 faulty bearing state and one fault-free state (data from different loads are denoted as one class) are contained in CWRU bearing dataset as shown in Table 3.2. While not every measurement contains data from three accelerometers - some measurements only have data from one acceleromter - the sample number in each bearing state is not equal.

3.3 Metalworking Fluid (MWF) datasets

Metalworking fluids (MWF) represent a range of oils or other liquids that can be used to cool, lubricate, and reduce friction during grounding, machining, and cutting process

Table 3.2: Case Western Reserve University bearing dataset [WJS22]

Bearing position	Bearing faulty part	Fault size	Motor load	Accelerometer position
DE	B	007	0, 1, 2, 3	BA, DE, FE
DE	B	014	0, 1, 2, 3	BA, DE, FE
DE	B	021	0, 1, 2, 3	BA, DE, FE
DE	B	028	0, 1, 2, 3	DE
DE	IR	007	0, 1, 2, 3	BA, DE, FE
DE	IR	014	0, 1, 2, 3	BA, DE, FE
DE	IR	021	0, 1, 2, 3	BA, DE, FE
DE	IR	028	0, 1, 2, 3	DE
DE	OR	007@3	0, 1, 2, 3	BA, DE, FE
DE	OR	007@6	0, 1, 2, 3	BA, DE, FE
DE	OR	007@12	0, 1, 2, 3	BA, DE, FE
DE	OR	014@6	0, 1, 2, 3	BA, DE, FE
DE	OR	021@3	0, 1, 2, 3	BA, DE, FE
DE	OR	021@6	0, 1, 2, 3	BA, DE, FE
DE	OR	021@12	0, 1, 2, 3	BA, DE, FE
FE	B	007	0, 1, 2, 3	BA, DE, FE
FE	B	014	0, 1, 2, 3	BA, DE, FE
FE	B	021	0, 1, 2, 3	BA, DE, FE
FE	IR	007	0, 1, 2, 3	BA, DE, FE
FE	IR	014	0, 1, 2, 3	BA, DE, FE
FE	IR	021	0, 1, 2, 3	BA, DE, FE
FE	OR	007@3	0, 1, 2, 3	BA, DE, FE
FE	OR	007@6	0, 1, 2, 3	BA, DE, FE
FE	OR	007@12	0, 1, 2, 3	BA, DE, FE
FE	OR	014@3	0, 1, 2, 3	BA, DE, FE
FE	OR	014@6	0	BA, DE, FE
FE	OR	021@3	1, 2, 3	BA, DE, FE
FE	OR	021@6	0	BA, DE, FE
Normal baseline			0, 1, 2, 3	DE, FE

[OJJ+22]. They play a significant role in metalworking processes. Firstly, MWF provide a layer of lubricant which acts as a cushion between workpiece and tool to reduce friction [McC94]. In addition, MWF reduce the heat and friction between tool and workpiece, and help to prevent burning and smoking [BKS+11]. Furthermore, MWF remove fines, chips, and swarfs from tool and workpiece surface [MJ08]. Therefore, they help to improve the quality of workpiece continuously. Finally, MWF protects workpiece and tool from corrosion when they are attacked by moisture, oxygen, acid substance, and dust [Smi08]. The MWFs are used in various metalworking processes such as grinding, cutting, drilling, and threading.

Generally, MWFs are categorized into water-based and oil-based fluids based on their basic substance [WBH12]. On the basis of water and oil, additives are added for MWF. The chemical additives of MWF affect their performance strongly. Even small changes of MWF additives can influence the performance of MWF in manufacturing processes significantly [WJDS22]. Over decades, more and more contributions focus on MWF additives and their performance in different metalworking processes. Most of these contributions evaluate MWF performance by measuring their pH value, acid split, bacteria, viscosity, sulphur, copper corrosion, residue, conductivity, foam, and surface tension [WPH17, Tra13, AP06]. Effect of MWF to workpiece and tool are measured by wear [HCMBS14], temperature [Can11], and taping torque test (TTT) [DS20]. However, only a few contributions concentrate on the MWF influence on workpiece and tool Acoustic Emission (AE) events.

Acoustic Emission is a passive non-destructive test (NDT) technique that makes use of the high-frequency acoustic energy emitted by an object undergoing stress [Lig21]. The cause of AE are attributed to the redistribution of the localized strain energy that is generated by the microscopic changes in a structure under different loading conditions [TA01]. Crack nucleation and growth, dislocation slips, grain boundary sliding, and phase transformation in structure lead to energy-releasing event [WJX+21]. According to [PLB+18], this energy-releasing event leads to a transient elastic wave, which will propagate in its medium and eventually reach to surface. The interaction of the transient elastic wave with a surface will cause surface motion. Special sensors which mounted on surface can pick the faintest surface motions and convert them into electrical signals, namely AE signals. Therefore, the characteristics of AE signal are: they reveal object energy naturally and manifest the dynamic processes in material [LBB+14]. As a consequence, AE signals are widely used to detect the presence of dangerous flaws and to locate the position of such flaws in various structure.

Thread forming is a manufacturing process that involves the generation of internal threading. According to [BFM+08], the formation of the thread is obtained by the successive action of the tap lobes. Each lobe causes a three-dimensional plastic flow. This plastic flow leads to an important strain hardening of the work material. Thread forming pushes bushing material in the flanks and cause a compaction of the micro structure by a chipless cold process [WDS21]. Because no chips affect the AE signals, the thread forming process maximizes the preservation of the original AE signal.

To obtain AE signals from different MWF in the process of thread forming, two laboratory experiments are conducted. These experiments are results from collaboration between the Chair of Dynamics and Control, University of Duisburg-Essen and Rhenus Lub GmbH & Co KG Mönchengladbach. Furthermore, measurements and data pre-processing are related to this cooperation.

The experimental test rig is shown in Figure 3.4 left. It consists of a tribometer of

type Tauro®120 (Taurox e. K., Germany), a test platform made of C45E (1.1191), a thread forming tool of the type Emuge M6-6HX InnoForm1-Z HSSE-TiN-T1 (as shown in Figure 3.5), different test and reference fluids, and a cleaning station with brushes and air blow system to remove chips and fluid residues. To avoid the influence of debris and chips on AE signals during these experiments, before testing, the test platform and the new tap are cleaned in an ultrasonic bath, dried in an oven at 50 °C, and cooled down to room temperature afterwards. Among different fluids, the tap is manually cleaned with a cleaning solvent. After each thread forming process, the tap is automatically cleaned in a cleaning station [WDS21].

Test platform has 368 (5.6H7 mm) pre-drilled holes of 28 mm in depth, arranged in 23 columns and 16 rows (from the back to front, the holes in the first column is named hole 1 to hole 16, the holes number in the second column are 17 to 32, the third column holes are named 33 to 48, etc.) [DS20]. For convenience, each thread forming process is named as one measurement and measurements in one column is named as one series [WLS22]. Test series are shown in Figure 3.4 right. The active tool length is 8 mm with a cutting lead of approximately 2-3 mm and a thread pitch of 1 mm.

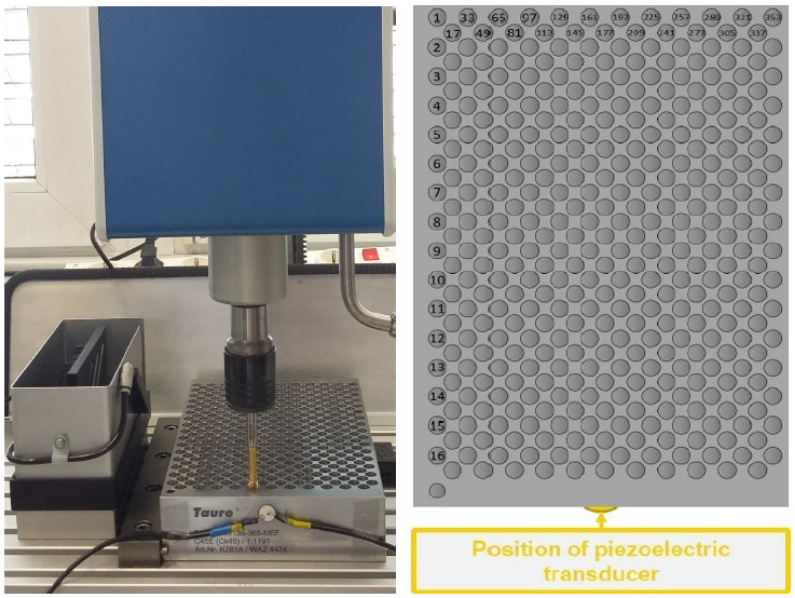

Figure 3.4: Test rig of MWF dataset [DS20]

A custom FPGA-based AE measurement system is used for the recording of the AE signals. At the front of the test platform, a disc-shaped broadband piezoelectric transducer is attached. The transducer is mounted using cyanoacrylic glue, has a diameter of 10 mm, a thickness of 0.55 mm, and a corresponding resonant frequency of 3.6 MHz. The AE measurements are acquired continuously at a sampling rate of 4 MHz [WDS21].

3.3.1 Dataset MWF19

In this experiment, eleven emulsion-based (reference fluid and other ten fluids) MWF are filled in pre-drilled holes. Before each test fluid, the reference fluid is applied to set same initial test conditions for each fluid. This means that the first column pre-drilled holes

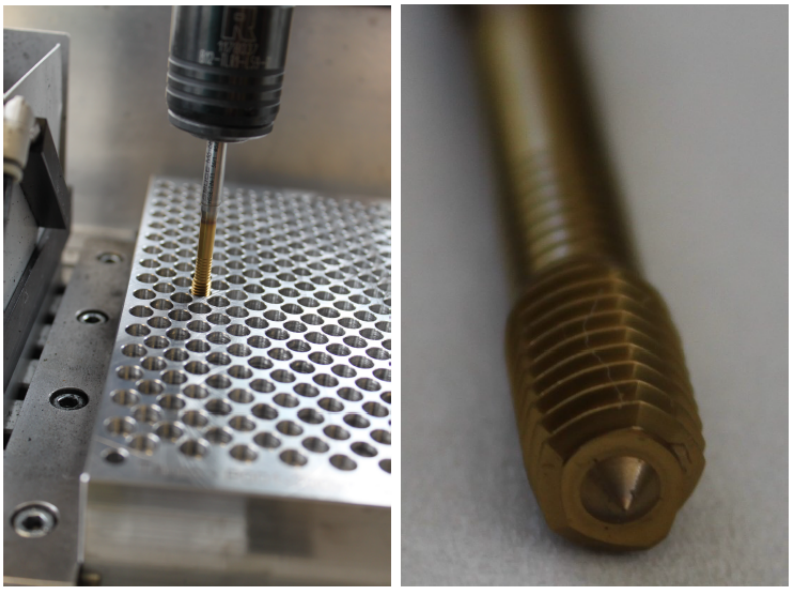

Figure 3.5: Thread forming tool of MWF dataset [DS20]

are filled with reference fluid, the second column pre-drilled holes are filled with fluid 1, the third column holes are with reference fluid again, the fourth column holes are filled with fluid 2, ... etc. In short, holes in the odd-numbered column are filled with reference fluid and holes in the even-numbered column are filled with the other ten liquids (series m01, m03, m05, ... are reference measurements while series m02, m04, m06, ... are test measurements) as shown in Table 3.3. As AE data from the last two columns are contaminated, so they are not considered in the calculation. Fluid and their additives that are applied in this experiment are listed in Table 3.4. Acoustic Emission signals taken from this experiment are stored in dataset MWF19 while this experiment is conducted in the year of 2019.

3.3.2 Dataset MWF16

In the second experiment, 112 threads of 28 mm in depth are formed at a speed of 1000 rpm using five different MWF. Besides the run-in of the tap at the beginning of the test procedure (32 threads with reference fluid), eight threads are tapped with each test fluid [WDS21]. Test series are shown in Table 3.5. In this experiment, both water-based (fluid 1 and fluid 2) and oil-based (fluid 3 and fluid 4) MWF are applied. The reference fluid is different from the reference fluid in the first experiment. Basis and additives of these five fluids are listed in Table 3.6. Acoustic Emission signals taken from this experiment are stored in dataset MWF16 as this experiment is conducted in the year of 2016.

Table 3.3: Test series in MWF19 [WJDS22]

Series	Threads	MWF typs
m01	1-16	Reference
m02	17-32	Fluid 1
m03	33-48	Reference
m04	49-64	Fluid 2
m05	65-80	Reference
m06	81-96	Fluid 3
m07	97-112	Reference
m08	113-128	Fluid 4
m09	129-144	Reference
m10	145-160	Fluid 5
m11	161-176	Reference
m12	177-192	Fluid 6
m13	193-208	Reference
m14	209-224	Fluid 7
m15	225-240	Reference
m16	241-256	Fluid 8
m17	257-272	Reference
m18	273-288	Fluid 9
m19	289-304	Reference
m20	305-320	Fluid 10
m21	321-336	Reference

Table 3.4: Metalworking fluid additives in MWF19 [WJDS22]

MWF	Additives	Additive substance
Reference	-	-
Fluid 1	Sodium sulfonate	4800 ppm
Fluid 2	Polysulfid, AS: Sulfur	1600 ppm
Fluid 3	Polysulfid, AS: Sulfur	2400 ppm
Fluid 4	Lauryl ethylene oxide phosphate	160 ppm
Fluid 5	Oleyl ethylene oxide phosphate	160 ppm
Fluid 6	Stearyl propylene oxide phosphate	86 ppm
Fluid 7	2-ethylhexylcocoate	8000 ppm
Fluid 8	Synthetic polymeric ester	8000 ppm
Fluid 9	Diethylene glycol	8000 ppm
Fluid 10	Polypropylene glycol	8000 ppm

Table 3.5: Test series in MWF16 [WJDS22]

Series	Threads	MWF typs
m01	1-32	Reference
m02	33-40	Emulsion 1
m03	33-48	Emulsion 2
m04	49-56	Oil 1
m05	57-64	Oil 2
m06	65-72	Reference
m07	73-80	Oil 2
m08	81-88	Oil 1
m09	89-96	Emulsion 2
m10	97-104	Emulsion 1
m11	105-112	Reference

Table 3.6: Metalworking fluid additives in MWF16 [WJDS22]

MWF	Basis	Water	Oil	Ester	Phosphorus
Reference	Water	95 %	0 %	1.25 %	50 ppm
Fluid 1	Water	95 %	1.4 %	0 %	3163 ppm
Fluid 2	Water	95 %	1.4 %	0 %	48 ppm
Fluid 3	Oil	0 %	85 %	6.5 %	80 ppm
Fluid 4	Oil	0 %	85 %	6.5 %	1600 ppm

4 State of art approaches

Machine learning is an area related to both cybernetics and computer science (or control science and computer science), attracting an overwhelming interest both of professional and of the general public [Fra20]. In the last few years due to successes of computer science (the emergence of GPUs, leading to significant improvements in the performance of computers and development of special software, allowing to work with big data), ML algorithms rise public interest significantly. Nowadays, ML has been applied widely in various fields. Overview of machine learning approaches is introduced in the first section.

Since the last century, IS has been researched for more than 70 years. A large number of experiments and questionnaire are conducted. Besides, considerable signals referring to IS and variety approaches are employed. Experiments, signals processing and classifiers are reviewed in the second section. The CWRU bearing dataset was open to public from 2012. As a benchmark, it is employed for validation of new designed approaches. Consequently, a large number of contributions are summarized and their results are compared. Although MWF datasets are not open to public, some approaches are still applied on them to discriminate MWF and their additives. Stat-of-art approaches that applied in IS, CWRU, MWF datasets will be summarized and reviewed in this chapter. In addition, results from these approaches will be shown and disadvantages with these approaches are summed up.

Parts of the contents, figures, and tables are prepared for publications of [WSS22], [WLS22] and [WJS22].

4.1 Overview of machine learning

Machine learning is a type of artificial intelligence that provides machines with the ability to automatically learn from data and past experience while identifying patterns to make predictions with minimal human intervention [DDPR15]. It derives insightful information from large volumes of data by leveraging algorithms to identify patterns and learn in an iterative process [Sun13]. In other words, ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model.

History of ML goes back to the 1943 with the first mathematical model of neural networks presented in the scientific paper 'A logical calculus of the ideas immanent in nervous activity' by [MP43]. In 1949, the book 'The Organization of Behavior' based the theories on how behavior relates to neural networks and brain activity and become one of the monumental pillar of ML development [Do49] . Psychologist Frank Rosenblatt also created a group that built a machine for recognizing the letters of the alphabet [Ros58] in 1950s. The first boom of ML happens at 1960s by deterministic approaches [Wid64, VL63] and stochastic approaches [MP69]. Machine learning development stagnation is in 1970s and 1980s. In the 1990s, machine learning work shifted from knowledge-driven approach to

data-driven approach [BM98, Man96, Mar91]. From the beginning of this century, it is ML rapid development time as the boost of big data, reducing cost of parallel computing and memory, and new algorithms of DL [WMZ09, Ayo10, Mit06, BN06, BB01]. In 2006, the term 'deep learning' was introduced by Hinton to explain new algorithms letting computer distinguish objects and text in images and videos [HOT06].

Learning process of ML algorithms can be divided into three parts according to [Chu18]:

i) Decision process: based on some input data (labeled or unlabeled), the algorithm produce a model from the input data.

ii) Error function: evaluate the performance of trained model.

iii) Model optimization: update weights autonomously until a metric threshold acquired.

Machine learning algorithms are molded on training data to create a model. when new input data are introduced to the trained model, the model uses the developed model to make a prediction. The performance of prediction is evaluated by metrics. Based on the metrics, the algorithm is either deployed or trained repeatedly until the desired metric value is achieved. Process of ML is shown in Figure 4.1. According to the development history

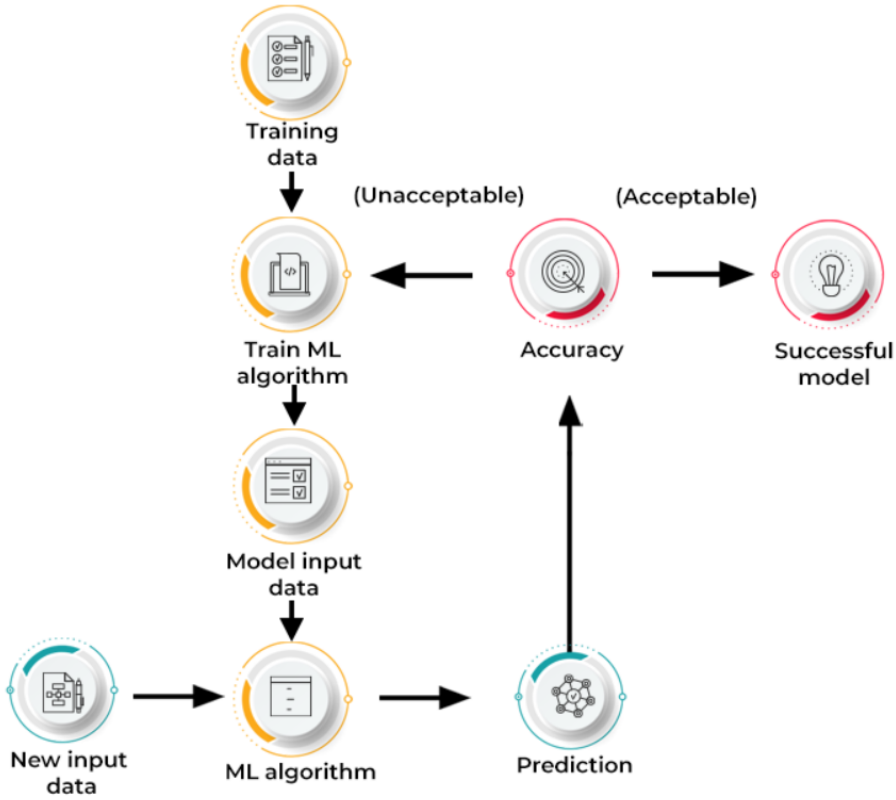

Figure 4.1: Machine learning workflow [Chu18]

and data processing difference, ML algorithms are divided into CML and DL. Conventional machine learning techniques are limited in capability of processing the data in their original form. These methods required considerable understanding and expertise for representation i.e. selection of features required professional knowledge strongly [Ong17]. In other words, CML is more dependent on human intervention to learn. Human experts determine the set of features to understand the differences between inputs, usually requiring more structured

data to learn [T+95]. On the contrary, DL eliminates some of data pre-processing steps that is typically involved in conventional machine learning. These algorithms can ingest and process unstructured data [KT18]. Deep learning algorithms determine which features are most important to distinguish by itself. Difference between CML and DL are shown in Figure 4.2. Based on whether input data are labeled or not, machine learning algorithms

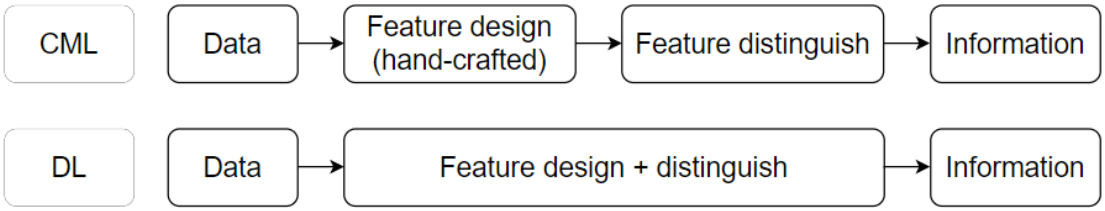

Figure 4.2: Difference procedure between CML and DL [KT18]

are divided into three primary categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms are trained on labeled data and usually for the task of classification and regression. The commonly used supervised learning methods are: Naive Bayes (NB), linear regression, logistic regression, support vector machine, K-nearest neighbor (K-NN), random forest (RF), decision trees (DT), Gaussian process regression (GPR), convolutional neural network (CNN), generative adversarial networks (GAN), self-organization maps (SOM). Unsupervised learning holds the advantage of being able to work with unlabeled data. The commonly used unsupervised learning methods are: K-means, hierarchical cluster, Gaussian mixture model (GMM), hidden mixture model (HMM), fuzzy C-means, autoencoder, deep belief network (DBN). Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. Reinforcement learning algorithms improve upon itself and learns from new situation using a trial-and-error method [QZ11].

Among various ML algorithms, neural networks are the heart DL algorithms. It is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operation [Mah20]. Neural networks architecture is made of individual units called neurons that mimic the biological behavior of the brain. Nowadays, many NN architecture have been validated and can be directly applied such as: LeNet-5, SqueezeNet, ENet, Dan Ciresan Net, VGG, AlexNet, Overfeat, and ResNet.

Due to its powerful functions, ML has been used in many fields such as: agriculture, image and speech recognition, traffic prediction, product recommendation, self driving cars, Email spam and malware filtering, virtual personal assistant, online fraud detection, stock market trading, medical diagnosis, and automatic language translation. To understand the principals ML apply on divergent field, review literature are summarized in this study. Large number of review contributions are available. [LBM+18] present a comprehensive review of research dedicated to applications of machine learning in agricultural production systems. An overview of past history, current developments, and emerging opportunities of ML for fluid mechanics are presented in [BNK20]. In [GKMJ22], authors provide readers with a gentle introduction to a few key machine learning techniques biological data. Maxwell et al. review the implementation of machine learning classification in remote sensing [MWF18]. Padarian et al. provide a comprehensive review of the application of ML

techniques in soil science aided by a ML algorithm to find patterns in a large collection of text corpora [PMM20]. Weichert et al. covers the majority of relevant literature from 2008 to 2018 dealing with machine learning and optimization approaches for product quality or process improvement in the manufacturing industry [WLS+19]. Qu et al. summarize the application of ML in microbiology [QGL+19].

4.2 State-of-art approaches on validating datasets

Inner speech has be studies for more than 70 years, many contributions relevant to IS are published. In this section, contributions that referring to IS study will be presented. Besides, the CWRU bearing dataset becomes a benchmark to verify diagnosis approaches since it was open to public in 2012. Approaches and their results referring to CWRU bearing dataset are also shown here. Although MWF datasets are not open to public by far, some contributions are still be involved. In this section, state-of-art approaches and relevant results of these three kinds of datasets are introduced.

4.2.1 State-of-art approaches on IS dataset

Since the second half of the 20th century, inner speech has been a research topic in philosophy and psychology [LHV18]. Varieties of Inner Speech Questionnaire (VISQ) has been used to link everyday phenomenology of inner speech – such as inner dialogue – to various psychopathological traits [ADMW+18]. Apart from questionnaire, a large number of experiments are conducted by meaning of inner speech to express phonemes, words, vowels, and phrases. Eight chronic post-stroke aphasic patients and thirteen cognitively healthy adults are underwent testing on a range of evaluative tests and four experiments designed to check whether chronic post-stroke patients has the ability to use inner speech [KGL+17]. Referring to the effect of overt speech on children's use of inner speech in short-term memory, three experiments are implemented in [HHSH91], meanwhile, the role of private speech and inner speech in planning during middle childhood is tested in [LMF10]. Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills is tested by in experiments in [LFMW09]. Direct evidence that inner speech sustains predictable task switching in adult are validated by experiment in [LMA+16]. University students are asked to listen to instrumental music and refer inner thoughts in a retrospective video-assisted interview to explore functions of inner speech and its expression in gesture [Fos20]. Neural correlates of inner speech and auditory verbal hallucinations experiments are held by Jones et.al [JF07]. In [WJP+17], two experiments are designed to compare electrophysiological signals between inner speech and overt speech. Lexical bias and the phonemic similarity effect in inner speech is validated by [BC09, Noo05]. To investigate the qualitative influence of inner speech on high and low measures of executive function, two experiments are conducted by [AB18]. Three experiments are carried out to examine the role of inner speech in task switching [EM03].

Patterns capturing brain waves relevant to inner speech can be divided into invasive and non-invasive. Since invasive data acquisition requires surgery operation, typically, this pattern is not popular for human. However, Stephanie et.al [MIM+18] acquire ECoG signals regarding inner speech from electrode grids, strips or depth electrodes that are temporarily implanted onto the cortical surface, either above or below the dura mater. Comparing with invasive pattern, non-invasive pattern is more friendly for subjects who

attend experiments. Through non-invasive experiments, MEG, fMRI, fNIRS, EEG, and sEEG signals are accessible. Magnetoencephalography signals with high magnetometers and gradiometers on inner speech are obtained by Dash et.al [DFBW21]. Neural activity during inner speech of meaningless syllable sequences was measured with MEG and fMIR from eight right-handed subject in [FHMO04]. Functional near infrared spectroscopy make use of electromagnetic rediation in near-infrared region in order to measure functional activation in cortical areas 1-3 cm beneath the scalp. Kamavuako et.al obtain fNIRS signals that referring to inner speech [KSG+18]. Comparing with ECoG, MEG, fMRI, and fNIRS signals, EEG signals are most widely used for inner speech analysis. Experiments gotten EEG data concerning inner speech data are conducted by [ZR15, CFC19, CGR17, NKA17, BDA21, CKFC20, GCTGRGVP17, SS14, HAMI17]. Both fNIRS and EEG data are obtained within eleven participants performing multiple iterations of three tasks by Rezazadeh et.al [RSYW+19]. Both EEG and fMRI data are recorded from ten healthy participants during covert speech in [YNB+16]. In addition, Angrick et.al record intracanial neural activity during speech processes using stereotactic electroencephalography (sEEG) electrodes [AOD+21].

Vowels, syllables, words, sentences, and states are performed by subjects and corresponding brain wave signals are acquired. To distinguish these signs, various traditional machine learning and deep learning approaches are applied. As the most classical machine learning method, SVM are utilized by [KSG+18, CKFC20, ZR15, CGR17, GCTGRGVP17] to distinguish inner speech data. Nguyen et.al [NKA17] use the variant of relevance vector machine (RVM) - variant of SVM - to distinguish signs in experiment. Random forest (RF) which establishes outcome based on the decision trees predictions is utilized by [CKFC20, CGR17, GCTGRGVP17] for inner speech distinction. As a classification and dimensionality reduction technique, linear discriminant analysis (LDA) is also used in [RSYW+19, AOD+21, CKFC20, SS14]. Because of its simplicity and efficiency, Naive Bayes (NB) is also applied for inner speech distinction [GCTGRGVP17]. K-Nearest Neighbor which is a non-parametric supervised learning classifier is also used for EEG data separation [HAMI17]. Another traditional machine learning method applied on fMRI and EEG data distinction is sparse logistic regression (SLR) [YNB+16]. Among deep learning, bidirectional long short-term memory recurrent neural network (BLSTM-RNN) which combine BLSTM and RNN are applied in [DFBW21]. Convolutional neural networks which is powerful for image distinction is used for inner speech classification by [CKFC20, CFC19]. Deep Belief Networks invented as a solution for the problems encountered when using traditional neural networks training in deep layered networks is also applied for inner speech discriminating [ZR15].

Following approaches designing in contribution above, performance of these approach should be accessed. Metrics for evaluating the performance of these approaches and classifiers are usually accuracy, stand deviation (SD), and F-score. Accuracy are most applied in the literature mentioned above. Besides, standard deviation is also employed to evaluate the results inequality among different subjects.For the imbalance classes, F-score can access the trained models unprejudiced. Results from part of contributions mentioned above are shown in Table 4.1.

Results from IS can be summarized as follows:

i) Most contributions apply their designed approach to the data from their own experiments;

ii) Words and vowels are the most applied signs in these experiments;

Table 4.1: Overview of inner speech studies [WSS22]

Data type	Sub.Num.	Classes	Classifier	Results	Reference.
fNIRS	8	6 words	SVM	ACC: 86.81, SD: 9.9	[KSG+18]
fNIRS+EEG	11	3 states	RLDA	ACC: 70.45 % SD: 19.19	[RSYW+19]
EEG	15	3 vowels, 5 words	RVM	ACC: 70.00 - 95.00 %	[NKA17]
EEG	10	4 words	EEG Net	ACC: 29.67 %	[BDA21]
EEG	15	5 vowels, 6 words	SVM	ACC: 18.71-22.25 %	[CKFC20]
EEG	15	5 vowels, 6 words	RF	ACC: 18.37-23.23 %	[CKFC20]
EEG	15	5 vowels, 6 words	RLDA	ACC: 20.77-26.66 %	[CKFC20]
EEG	15	5 vowels, 6 words	CNN	ACC: 24.35-29.58 %	[CKFC20]
EEG	15	5 vowels, 6 words	EEGNet	ACC: 24.46-30.25 %	[CKFC20]
EEG	12	7 syllabic, 4 words	SVM	ACC: 18.08-79.16 %	[ZR15]
EEG	12	7 syllabic, 4 words	DBN	ACC: 80.00-91.00 %	[ZR15]
EEG	15	5 vowels	CNN	F-score: 33.17-36.65 %	[CFC19]
EEG	15	5 vowels, 6 words	SVM	ACC: 18.26-21.94 %	[CGR17]
EEG	15	5 vowels, 6 words	RF	ACC: 19.60-22.72 %	[CGR17]
EEG	27	5 words	RF/NB/SVM	best ACC: 58.41 % SD: 12.41	[GCTGRGVP17]
EEG	4	2 words	K-NN	ACC: 58.00 %	[HAMI17]

iii) Most results accuracy are under 50 % and accuracy which reach up to 90 % only come from a few contributions;

iv) Standard deviation of some approaches are very high;

v) Only one contribution ([BDA21]) apply the same dataset as in this study. Accuracy from it is very low: only 29.67 %.

4.2.2 State-of-art approaches on CWRU dataset

Large number of new diagnosis approaches are designed for rotating machinery as its importance in industry in the past years. Many contributions verify their approaches performance with CWRU bearing dataset. A large number of CML are applied to CWRU bearing dataset. K- nearest neighbors are employed as a classifier to differentiate bearing states in [WWW$^+$21, LWSH19, RABV14, VSP$^+$22]. As a traditional classifier, SVM is also applied on CWRU bearing dataset [SAF15, YP19, WCJ17, ZPC17]. [SA18, PBMT21] employ Naive Bayes to differentiate CWRU bearing states. Decision trees is employed for classify CWRU bearing states by [ZYZ21]. [KMC$^+$21, VGK17, WP19, RDC20, LLL18] put random forest into use for CWRU data classification. Besides these supervised learning methods, numerous unsupervised learning approaches are also utilized. K-means is applied for differentiating CWRU bearing states in [ZNZ$^+$16, ZZL$^+$17, WZLL21, ZLT$^+$20]. Gaussian mixture model are also employed to differentiate bearing states in CWRU bearing dataset by [Yu11, ARMH20, AAL$^+$21, LZLL15]. In [WHW21, PWC$^+$21, FLXL16], fuzzy C-means is employed to discriminate bearing patterns.

In addition to CML, numerous DL approaches are referred to CWRU bearing dataset. As a sophisticated generative model that employs a deep architecture, deep belief network is applied for bearing states discrimination in [WLR$^+$15, LWDZ18, YH17, PCZ18, SJWW17]. Convolutional neural network which has become dominant in various computer vision tasks is also widely applied for CWRU bearing dataset classification [WLGZ17, QWW$^+$19, LLT$^+$17, XLX$^+$17, LVD21]. To learn a lower dimensional representation for a higher dimensional data, autoencoder is applied to capture the most important characters of vibration signals in [YFCZ18, WLGZ17, SJLL18, DW$^+$18, XT19, DTS$^+$19, YWLZ19]. Generative adversarial networks use two neural networks, pitting one against the other in order to generate new, synthetic instances of data that can pass for the real data. Besides its widely application in image, video, and voice generation, it is also applied for bearing distinction [Hua19, JHZ$^+$19, ZCL$^+$20], As a special type of artificial neural network adapted to work for time series data or data that involves sequences, recurrent neural networks is also applied for CWRU bearing states distinction in [ZZH$^+$21].

Compared with CML and DL, TL offers greater flexibility in extracting high-level features transferred from the source to the target problem. Contributions [ZLLN18, ZLC$^+$19, CLZG20] prove that TL is a powerful algorithm for CWRU bearing states analysis. Review of these CML, DL, and TL approaches and their results are shown in Table 4.2.

From Table 4.2, the following can be concluded: The results from other contributions can be summarized as:

i) Results from these contributions are good. All accuracies are higher than 90 %;

ii) Accuracies from some contributions are very good. Results from some contributions [ZLLN18, WLGZ17, WLR$^+$15, XLX$^+$17, WLGZ17, DTS$^+$19, YWLZ19, PCZ18, YFCZ18, ZYZ21, ZCL$^+$20] are over 99 % ;

iii) Some approaches get perfect results for data they selected via corresponding approaches. Accuracies from contributions [YH17, WP19, RDC20, PWC+21] reach values up to 100 %.

However, inadequacy of these approaches are also obvious:

i) Selected data that used in relevant approaches are not clear in some contributions [SJLL18, WLGZ17, LWDZ18, PCZ18, ZNZ+16, ZZL+17, WZLL21, WHW21, PWC+21]. While data are the most critical part in machine learning, results are not convincing when data used are not pointed out [WJS22].

ii) For some contributions, only part of the data are selected - data from drive side faulty bearing and baseline are used in [ZLLN18, YP19, WLR+15, SJWW17, YH17, YFCZ18, LLL18, LZLL15, ZCL+20]. Data from drive end faulty bearing and baseline are easier to differentiate, good results from easily distinguishable data can not demonstrate the robustness of these approaches.

iii) Most study classify selected data into 4 or 10 classes. Four classes denoted as fault-free, fault at inner race, ball, and out race while ten classes denoted as fault-free, fault location in different parts (inner race, ball, out race) combing with fault size (7, 14, 21 mils). Fault size of 28 mils and different fault locations in out race can not be distinguished in four or ten classes [WJS22].

iv) In some publications only a small amount of information is referred on data processing, parameters, and hyperparameters optimization. If only parts of well-performed samples are taken into calculation, results are unconvincing [WJS22].

v) Accuracy is mostly applied to evaluate the performance of these approaches. When sample numbers in each class are not equal, accuracy is not suitable to evaluate approaches and other metrics are needed to judge approaches performance like F-score [WJS22].

4.2.3 State-of-art approaches on MWF datasets

Metalworking fluid experiments were conducted by the collaboration between the Chair of Dynamics and Control, University of Duisburg-Essen and Rhenus Lub GmbH & Co KG Mönchengladbach. Datasets from these two experiments (MWF19 and MWF16) are still not open to public. However, Mr. Wirtz and Miss Demmerling applied an approach which combining with CWT and K-means to them [WLS22]. In their approach, AE signals are transformed from time domain to time-frequency domain by CWT firstly. Different process phases - forward and reverse - can be distinguished according to peak frequencies of AE signals. Furthermore, K-means clustering is employed to differentiate the scalogram.

Results for MWF19 from the approach is shown in Figure 4.4 [DWS22] .

Results for MWF16 from the approach is shown in Figure 4.5 [WDS17]. From results of both datasets, such conclusion can be drawn:

i) Approaches applied in these two datasets are monotonous. Data in both two datasets are distinguished by the approach combining CWT and K-means [WLS22].

ii) Data distinguish just come from partly series: In MWF2016, data from series 2 and 3; data from series 9 and 10; data from series 4 and 5; data from series 2, 4 and 11 are compared. In MWF2019, data from series that used similar MWF (series 4 and 6; series 8, 10, and 12; series 14 and 16; series 18 and 20) are distinguished.

iii) Clustering results are not ideal, especially in MWF2019, results from series 7 and 8 are only 40 %; results from series 4, 5 and 6 are only 41 %.

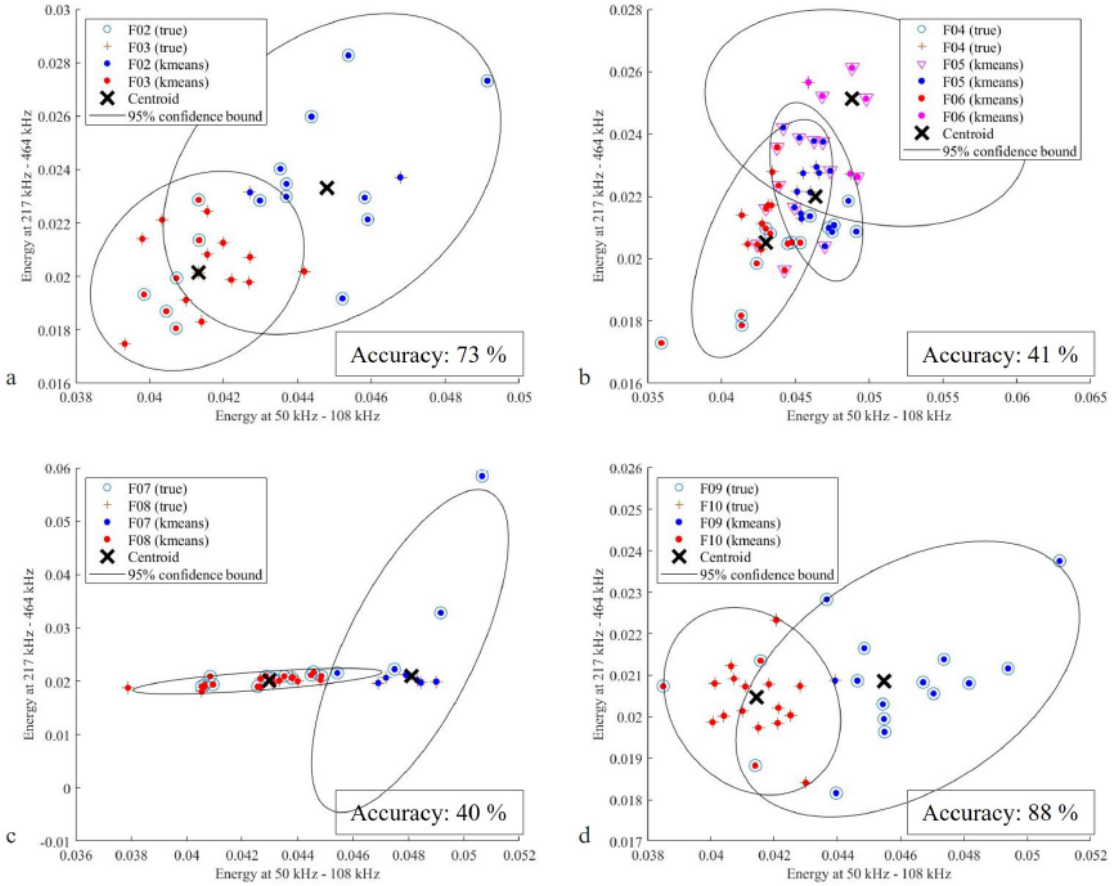

Figure 4.3: Results for dataset MWF19 [DWS22]

iv) No approach is used for differentiate all MWF from one dataset in [DWS22], [WDS17], just part series are distinguished in one dataset.

v) Although both datasets store AE signals from MWF, it is still ambiguous whether TL can be applied to these two datasets.

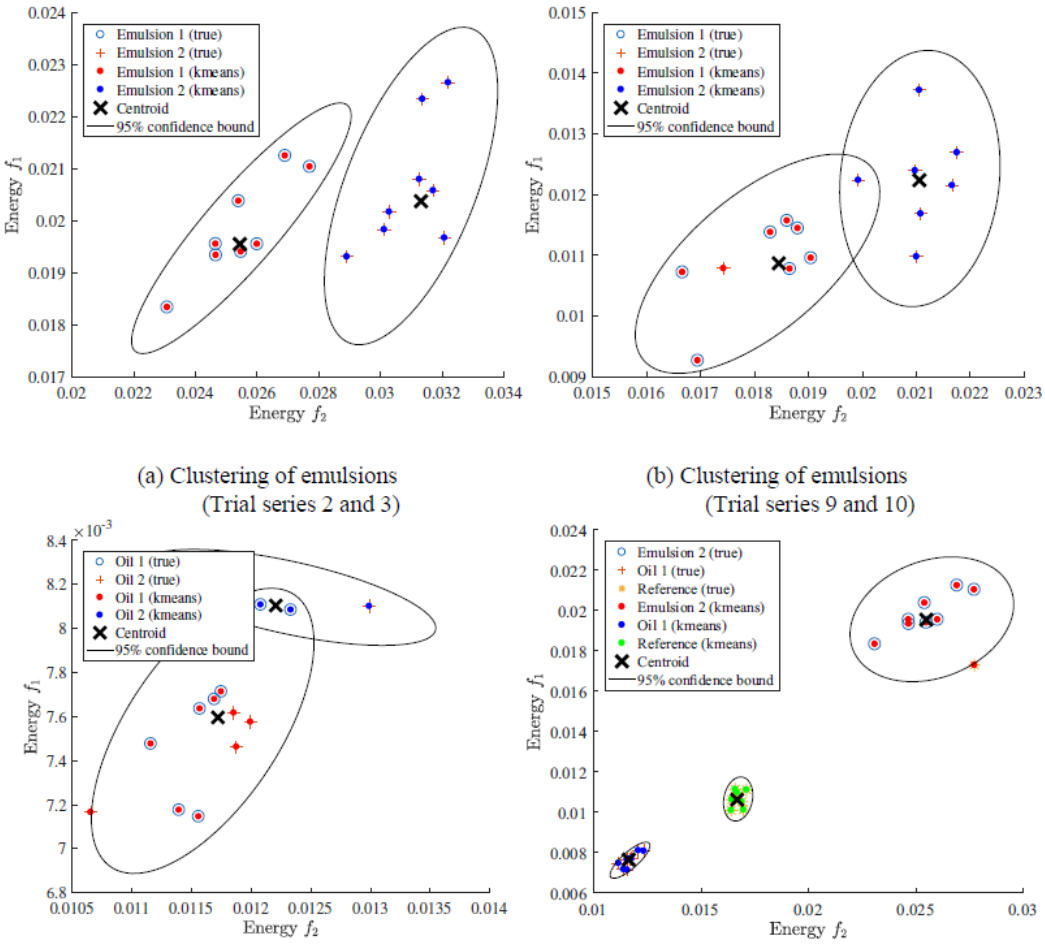

Figure 4.4: Results for dataset MWF16 [WDS17]

Table 4.2: Results comparison of CWRU dataset with different approaches [WJS22]

No.	Selected data	Fault size (mil)	Fault part	Classes	Classifier	Accuracy	Reference
1	–	7, 14, 21, 28	IR, B, OR@3,6	12	Autoencoder	97.18	[SJLL18]
2	DE, 12k	7, 14, 21	IR, B, OR@6	10	CNN	99.6	[ZLLN18]
3	–	7, 14, 21	IR, B, OR	10	CNN	99.79	[WLGZ17]
4	DE, 12k	7, 14, 21, 28	IR, B, OR@6	4	SVM	96.98	[YP19]
5	DE, 12k	7, 14, 21	IR, B, OR	10	DBN	99.55	[WLR+15]
6	DE, 12k	7, 14, 21	IR, B, OR@3,6,12	16	DBN	98.75	[SJWW17]
7	-, 12k	7, 14, 21	IR, B, OR	10	CNN	97.97-99.75	[QWW+19]
8	-, 12k	7, 14, 21	IR, B, OR	10	CNN	99.41	[XLX+17]
9	-, 12k	7, 14, 21	IR, B, OR	10	TL	99.82	[WLGZ17]
10	-, 12k	7, 14, 21	IR, B, OR	4	Autoencoder	99.6	[DTS+19]
11	-, 12k	7, 14, 21	IR, B, OR	10	Autoencoder	99.73	[YWLZ19]
12	–	7, 14, 21	IR, B, OR	10	CNN	96.95	[LWDZ18]
13	DE, 12k	7, 14, 21	IR, B, OR@3,6,12	10	DBN	100	[YH17]
14	–	7	IR, B, OR	4	DBN	99.58	[PCZ18]
15	DE,-	7, 14, 21	IR, B, OR	10	Autoencoder	99.6	[YFCZ18]
16	-, 12k	7, 14, 21	IR, B, OR@3,6,12	11	Naive Bayes	98.27	[PBMT21]
17	-, 12k	21	IR, B, OR	4	DTs	99.2	[ZYZ21]
18	-, 12k	7	IR, B, OR	4	Random forest	100	[WP19]
19	-, 12k, 48k	7, 14, 21	IR, B, OR	2	Random forest	100	[RDC20]
20	DE, 12k	7, 14, 21	IR, B, OR	10	Random forest	96.75	[LLL18]
21	–	14, 21	IR, B, OR	4	K-means	90.63-93.13	[ZNZ+16]
22	–	–	IR, B, OR	4	K-means	93.13	[ZZL+17]
23	–	7, 14, 21	IR, B, OR	4	K-means	97.73-97.99	[WZLL21]
24	DE, 12k	7, 14, 21, 28	–	5	GMM	97.2	[LZLL15]
25	–	7, 14, 21	IR, B, OR	10	FCM	95	[WHW21]
26	–	7	IR, B, OR	4	FCM	100	[PWC+21]
27	DE, 12k	7	IR, B, OR	4	GAN	99.2	[ZCL+20]

5 Proposed approaches

Except for some public databases that have been processed by the designer, most practical data and datasets contain some kind heterogenity or contamination, which called 'outliers' [Kay04]. Outliers have a disproportionate effect on statistical results which result in misleading interpretations. Isolated outlier has positive impact on the results of data analysis and data mining. Furthermore, to obtain accurate data in many experiments, a lot of foreshadowing or extension steps are also required in experiments. However, data from these foreshadowing or extension steps are not required and they need to be removed. In addition, large databases computation requires very powerful computer configuration and there are time consuming. These computer configurations are not available in every laboratory. To reduce computer configuration requirements and computing time, data selection is also necessary.

Conventional machine learning techniques have restricted ability on processing data from their original form and they require considerable understanding and expertise for features selection. Although DL algorithms can ingest and process unstructured data and determine which features are most important to distinguish, however, for most new DL approaches data processing are also needed to improve models ability in data distinguishing. In the step of data processing, many techniques can be applied such as: data augmentation, segmentation, transformation, and normalization. There is no unified standard on which data processing method is more suitable for all datasets. Appropriate data processing methods can only be selected based on experience or professional knowledge. In proposed approaches, except for data processing methods that selected based on experience, other data processing methods are also employed.

According to the destination of data distinguish, various ML methods can be applied. SVM, NB, CNN, K-NN are employed for the task of classification; linear regression, SVR, GPR are applied for the task of data regression; GMM, k-mean, hierarchy cluster, hidden mixture models are useful for the goal of data clustering. For a certain task, no conclusion to the most acceptable algorithms currently. Among some contributions, different ML approaches are compared when specific to one dataset. In proposed approaches, besides ML approaches that selected by experience and literature review, other ML approaches are also tried out.

For machine learning algorithms, hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up [DSW+18]. Values of hyperparameters have a significant impact on results of ML. Parameters in data processing methods also effect ML results. Consequently, hyperparameters in machine learning and parameters in data processing need to be tuned or optimized. Less information on how parameters and hyperparameters are tuned in most contributions, in this study, parameters and hyperparameters optimizing process will be presented in detail.

Usually, data selection, data processing, features differentiating, and parameters and hyperparameters optimization are included for most ML algorithms. For a few basic datasets and benchmark like MINIST, CIFAR 10, CIFAR 100, ImageNet, as they have been processed by designers, no more data processing steps are needed. Flowchart of ML is shown in Figure 5.1.

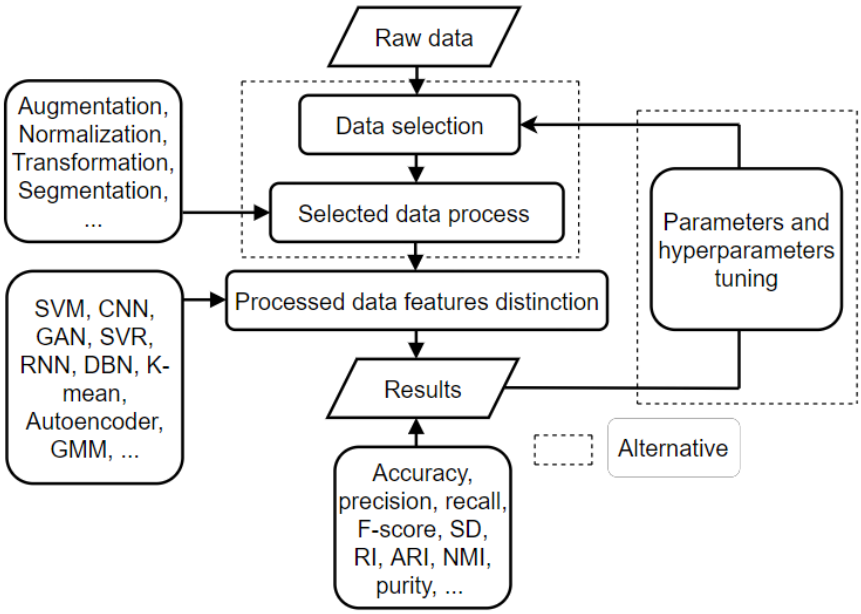

Figure 5.1: Flowchart of machine learning

Based on the flowchart of ML, according to the focus difference in each step, three big categories approaches which can be divided into five sub categories are proposed in this study. The first approach is a supervised learning algorithm focus on ML hyperparameters optimization and data selection, less other data processing methods are applied in this approach. Approach 2 is also supervised learning. In approach 2, more data processing methods, parameters and hyperparameters optimization algorithms are applied. Based on the approach designed in approach 2, a transfer learning (approach 2.0) is raised between MWF19 and MWF16. In the third approach, diversity data are chosen, various data processing methods, and features distinguish algorithms are integrated. Both supervised learning (approach 3.1) and unsupervised learning (approach 3.2) methods are applied on the approach 3. Detail of these approaches will be introduced in the following section.

The contents, figures, and tables presented in this chapter are modified after previous publications [WDS21], [WS22], and [WJDS22]. Part of the contents, figures, and tables are prepared for publications of [WSS22], [WJS22], [WLS22], and [DWS22].

5.1 Approach 1: Focus on ML hyperparameters tuning

A model is defined by the model hyperparameters in ML. Hyperparameters chosen has significant impact in ML training process as they map input features to the targets. The process of training a model involves choosing the optimal hyperparameter [WDS21]. For

different ML algorithms, various hyperparameters should be tuned. Train-test split ratio, learning rate in optimization algorithm, choice of activation function, choice of cost or loss function, number of iterations, layers structure in neural network, batch size, etc... are hyperparameters need to be tuned. Approach 1 focus on ML hyperparameters tuning and a few steps on data processing. Classifier employed in this approach is CNN, therefore, hyperparameters referring to CNN structure and hyperparamters relevant to training algorithms will be tuned according to each dataset. For data processing, a new data selection method is proposed in this approach. After proper data are selected, only data segmentation is applied in the step of data processing. Workflow of this approach is shown in Figure 5.2.

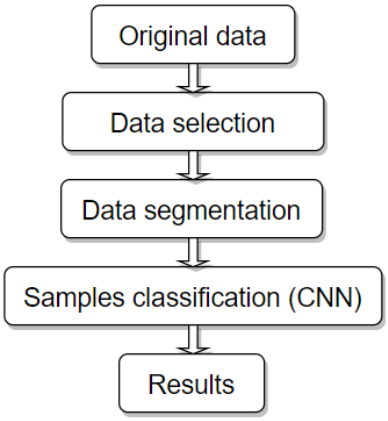

Figure 5.2: Workflow of approach 1

This approach is applied to CWRU bearing dataset and MWF datasets. Since CWRU bearing dataset is a benchmark and data has been preliminary processed by CWRU school of engineering, consequently, it is not complicated for data selection - data with 12k sampling rate are all chose. Most measurements contain data from three channels while some measurements contain data only from one channel from Table 3.2. Channels difference among measurements is shown in Figure 5.3. In the process of classifying faulty free and fault bearings from drive end into ten classes, both three channels data and single channel (DE) are employed. To distinguish all 29 bearing states, only DE channel data are employed as all bearing states measurements contain DE channel data. Fan end channel data and baseline data are not concluded in all measurements.

Unlike CWRU bearing dataset, only one accelerometer is attached into the work piece, therefore, only all measurements contains data from one channel. Although more test series on reference fluid than other fluids in both experiments, part of reference test series are applied to ensure all classes contain same number of samples. In the experiment of MWF19, data from the last two series are damaged, so they cannot be employed. Besides, eleven series measurements are from reference fluid while one series measurements for each test fluid. To reduce the influence of data imbalance on the results, measurements number for all kind of MWF should be the same. Therefore, although eleven series measurements comes from reference fluid, data in series m01 are selected [WDS21]. In addition, as each measurement is around 5 seconds with sample rate of 4 MHz, as a result, about 20 million data are included in one measurement. Consequently, it is impossible to calculate data

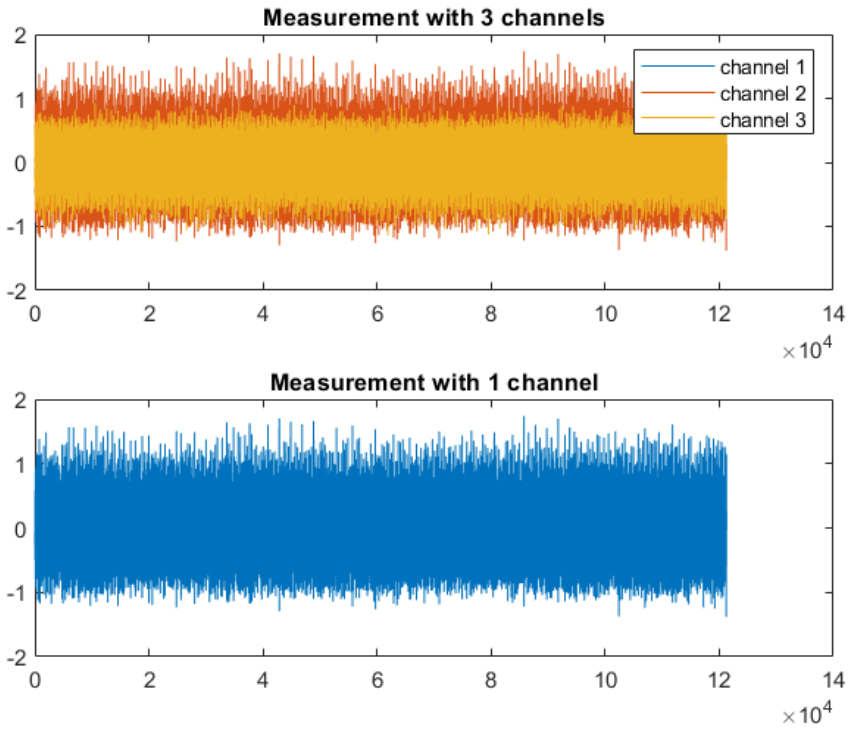

Figure 5.3: Measurements of CWRU with various channels

from all measurements in each series. From physical point of view, when measurements locations are far away from the piezoelectric transducer, the AE signals are weak. When measurements locations are very near to the piezoelectric transducer, the AE signals could contain much noise. Consequently, although 16 measurements are in each series, 4 measurements which have almost the same distance to piezoelectric transducer in each series are chosen. Considering measurements locations influence on AE signals, data from measurement 7 to measurement 10 in each series are selected [WDS21]. In the experiment of MWF16, 32 measurement are tested in run-in procedure. Besides the run-in of the tap at the beginning of the test procedure, eight threads are tapped with each test fluid. From series m02 to series m11, each series contains eight threads, in order to reduce the impact of hole location, the middle four threading AE signals are used. For example in series m02, the position of the threads is from measurements 33 to 40. For data processing, only the data from measurements 35 to 38 are taken into account. In order to balance the samples number in every class, different number of threads are applied in m01, all these threads are chosen from the middle of series m01.

Another side in data selection should be considered is the different phases in thread forming. Thread forming process is divided into forward and reverse phases theoretically. However, thread forming process in this experiment are conducted manually, air phases are unavoidable included in each measurement. From physical point of view, no usable AE data in air part as the tap has no contact with the platform [WJDS22]. For this reason, data in this part should be removed. Threads are mainly formed in the forward part of one measurement, therefore the relevant AE events mainly occur in this part. As a results,

data in this part should be analyzed. However, no clear boundary among air, forward and reverse part in the experiments of MWF16 as shown in Figure 5.4.

To find the boundary among different part in one measurement and pick up the forward part data, a new method to distinguish different parts are suggested. Firstly, whole measurements are transformed into time-frequency domain by CWT and scalogram are gotten. As shown in Figure 5.5, boundary among different parts are clear in scalogram, forward part can be isolated. While the forward part of each measurement is clear in scalogram in time axis, the forward data in time domain can be simply selected. As the sampling rate of AE signals is 4 MHz, start time multiply with 4,000,000 is the start point of forward data [WS22]. Meanwhile, end time multiply with 4,000,000 is the last data point of forward part. Data in between start point and end point are forward part data. Despite the start time of each threading is different, forward part duration time are the same for all threading. Comparing with dataset MWF16, as shown in Figure 5.4, boundaries among different parts in raw AE signals in MWF2019 are very clear.

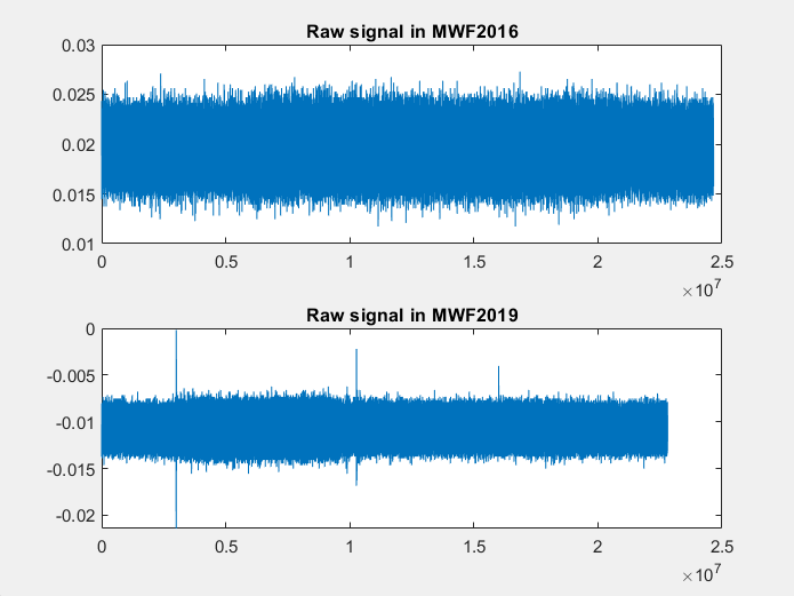

Figure 5.4: Raw AE signals. up: signal in MWF2016; down: signal in MWF2019

Since samples are at the core of machine learning, a large amount of training samples plays a critical role in making the machine learning models successful. To train a machine learning model, the sample number must be pretty large. Nevertheless, there are 308 measurements in CWRU bearing dataset, 112 measurements in MWF16 dataset, and 336 measurements in MWF19 dataset in total. Such small measurements number definitely cause deviations in the calculation results. So data augmentation is necessary. Measurement segmentation is an efficient technique for increasing samples number. While data used in these datasets are all from rotating machinery, data within one rotation cycle is important for rotating machinery, segment length can be decided according to rotating speed. Data in one rotation cycle can be determined by sampling rate f_s and rotating speed rpm. Calculation of data in each rotation can be calculated by the following equation:

$$N_{\text{spr}} = \frac{f_s}{rpm} * 60 \tag{5.1}$$

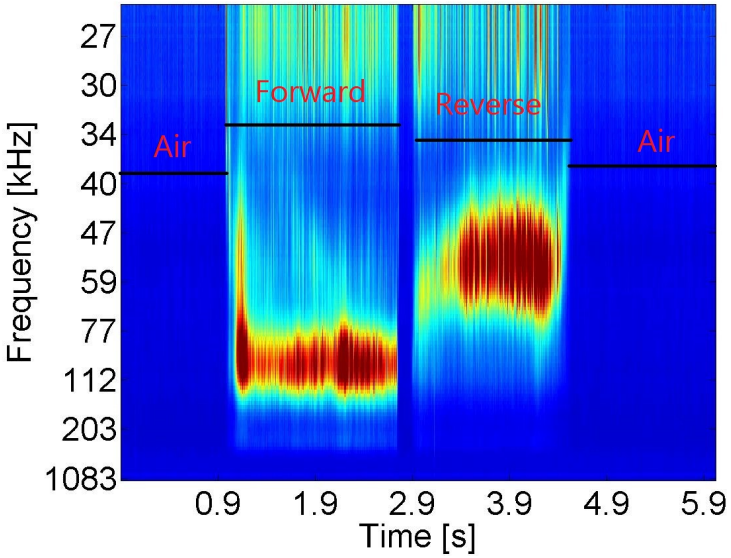

Figure 5.5: Scalogram of AE singal [WS22]

with

N_{spr} : number of samples per round,

f_s : sample rate, and

rpm : rounds per minute.

For the CWRU bearing dataset, as the motor speed variate from 1720 to 1797 rpm, the number of samples per round is not fixed under different load. Considering Equation 5.1 and a sample rate f_s with 12 kHz, the number of samples per round differ from about 419 to 400. Average of data in each rotation is 408. To diminish important data loss during measurements segmentation, adjacent segments are overlapped. The overlap among neighboring segments are 0.5. After segmentation, samples can be seen in Figure 5.6 Rotating speed is singular for MWF datasets, therefore, segments length is unify. In the experiment of MWF19, the speed is 1061 rpm, 226200 data in each rotation correspondingly. In the experiment MWF16, speed of tool is 1000 rpm, correspondingly, 2400000 data in each rotation. Like the segment process in CWRU bearing dataset, adjacent segments are also overlapped with 0.5 overlap rate. Segments are shown in Figure 5.7. After measurements are segmented into samples, the samples number increasing so greatly that they can be put into CNN for classification. Structure of CNN and hyperparameters referring to training algorithms are designed according to each dataset. Besides, the CNN structure is also designed according to the functionalities of different layers and activation functions (as introduced in chapter 2). Based on these two points, as less data in each sample for CWRU dataset, less convolution layers are applied. On the contrary, complicated structure are applied for MWF dataset as more data in each sample. Structure of CNN for each dataset are presented in Figure 5.8. Besides hyperparameters relevant to CNN structure, hyperparameters relevant to training algorithms like initial learning rate, batch size, maximal number of epochs, and drop probability are also tuned separately and

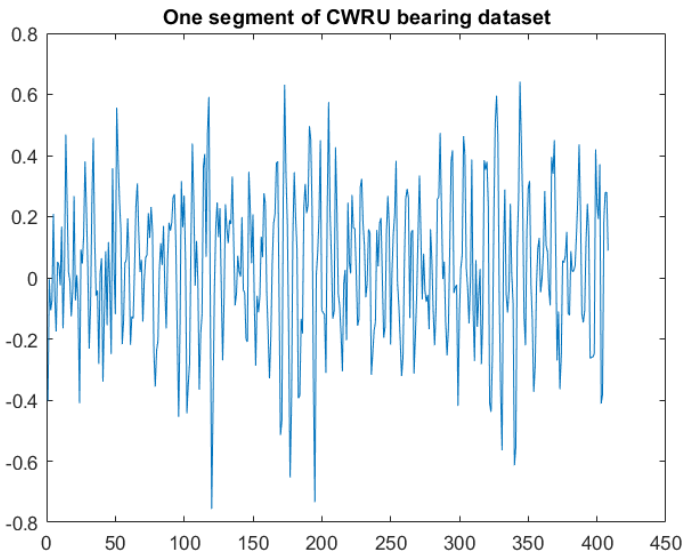

Figure 5.6: Segment of CWRU bearing dataset

manually. In the process of hyperparameters referring to training algorithms, a method called 'cross validation' is applied. 'Cross validation' means that each hyperparameter are tuned separately and then combination of best performed values are applied to the model. For example, various values are pointed to initial learning rate, the best value of this hyperparameter can be settled down according to the results. Afterwards, best drop probability value is also determined by the same way. Lastly, all the best values are combined together to get the final results.

Comparing with approaches from other literature, highlight of this approach is as follows:

i) More data are applied in this approach comparing with other contributions in CWRU bearing dataset.

ii) A new method is proposed for searching for bonds among different phases in MWF16 dataset. Continuous wavelet transform is used to find boundaries of different parts in data selection step for MWF datasets. As CWT is often used for time-frequency analysis in the past, applying it for bounds distinction is also a new point in this approach.

iii) Segments length is decided according to the rotating speed. Comparing with other contributions, this is also innovation. As although some other contributions also used the data segmentation technique, there is less information on the way they segment measurements and reasons they segment measurements.

iv) Few data processing methods are applied for feature extraction, only measurement segmentation is employed in data processing after data selection.

v) Convolution neural network structure are designed pinpoint to each dataset.

vi) Hyperparameters relevant to training algorithms on CNN are tuned by 'cross validation'. In other contributions, there is few information on how the hyerparameters in ML are tuned.

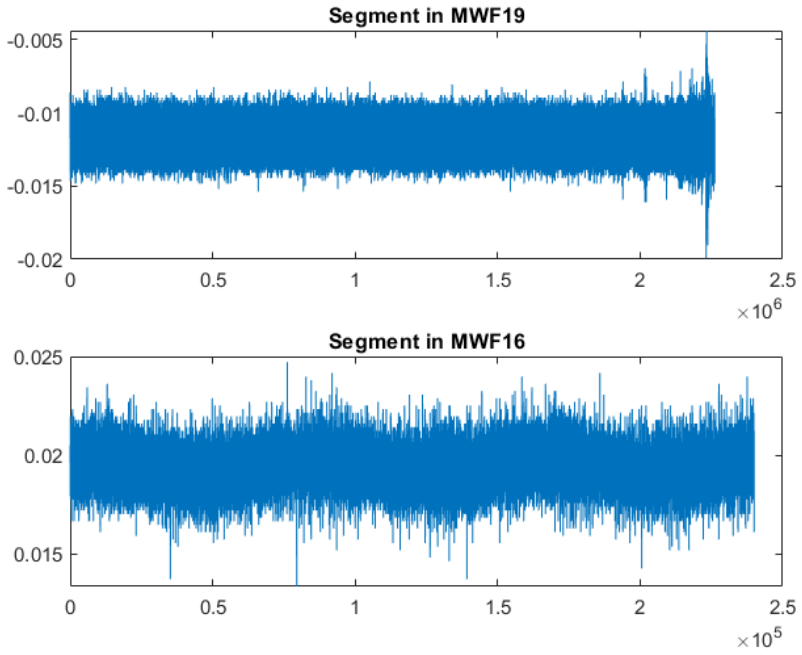

Figure 5.7: Segment of MWF datasets [WDS21]

5.2 Approach 2: Emphasis on data processing, parameters, and hyperparameters optimization

As introduced in previous section, segment length is fixed, less steps in data processing and hyperparameters are tuned manually in approach 1. Unlike approach 1, segment size is settle into various values and more data processing steps are included in approach 2. Besides, parameters in data processing and hyperparameters in CNN are tuned automatically. Furthermore, parameters and hyperparameters are optimized in one step as one objective function. Moreover, as the results from MWF19 are very good, MWF16 dataset and MWF19 dataset are very similar with each other, a transfer learning is proposed between two MWF datasets. Parameters and hyperparameters are transferred from MWF19 to MWF16.

5.2.1 Overview of approach 2

Approach 2 applied for CWRU and MWF datasets is more complicated than approach 1. Besides data selection and segmentation, segments are transformed, data and spectrogram are normalized before samples are classified by CNN. In addition, parameters in data processing and hyperparameters in CNN are optimized together. A loop from data segmentation to CNN is build [WJS22]. Details information is as follows: firstly, data are selected as in approach 1. Afterwards, selected measurements are segmented into different segments to increase sample number and balance sample number of each class. Furthermore, segments are transformed from time domain to time-frequency domain by STFT and spectrogram are gotten. To reduce bad results caused by outliers, both segments and spectrogram are normalized. Lastly, normalized spectrogram features are extracted

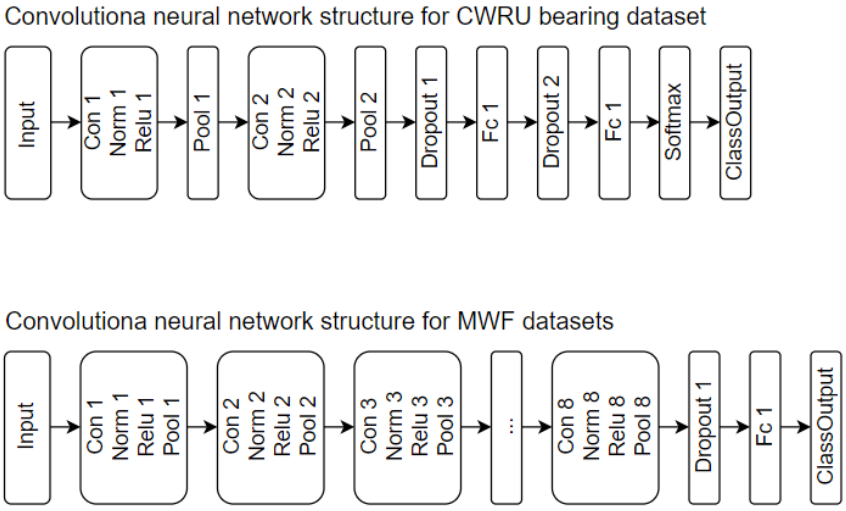

Figure 5.8: Convolution neural network structure of approach 1 [WDS21]

and classified by CNN. Additionally, parameters and hyperparamters are optimized automatically in one step [WJS22]. Flowchart of proposed approach is shown in Figure 5.9. Besides the data selection method in approach 1, highlight of approach 2 is as follows:

i) More data processing steps are included to help search data features.

ii) Instead of identical segments length, various segments length are tried.

iii) To balance sample number in each class for CWRU bearing dataset, overlap among segments in each class is various.

iv) Segments and spectrograms are normalized to remove the outliers.

iv) Parameters in data processing and hyperparameters in CNN are optimized together as one objective function.

v) A loop is formed among data segmentation, transform, normalization, and classification.

Since classification results are very good for 10 classes by approach 1. All 29 bearing states in CWRU bearing dataset are classified in this approach which means that all data with 12 kHz sampling rate are applied in this approach. In the step of measurements segmentation, segment length is not identical. For CWRU bearing dataset, for different bearing fault size and location, The measurement numbers are not identical as shown in Figure 5.10. In most measurements, data come from 3 channels (BA, DE, and FE) and under 4 motor loads (0, 1, 2, and 3), that is 12 measurements. But in some cases data acquisition are only taken from one channel or one motor load, in other words, in this class only 3 or 4 measurements data could be used.

Test time for each measurement is different in CWRU bearing dataset. With all data and classes from Figure 5.10 and a fixed segment size, this leads to the segments in each class are not identical. To balance the distribution, segments in each class is divided with an individual percentage of overlapping. The individual overlap percentage is chosen so that an equal number of segments per class N_s is achieved. The individual overlap percent of each class O_i is calculated by

$$O_i = 1 - \frac{N_i * T_i * f_s}{N_{\text{sps}} * N_s} \tag{5.2}$$

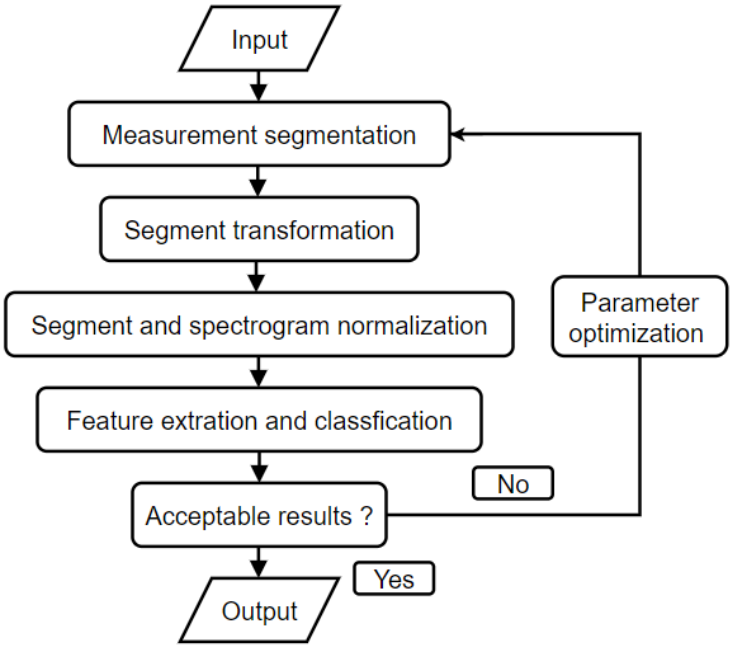

Figure 5.9: Workflow of approach 2 [WJS22]

with

N_i : number of measurements per class,

T_i : mean of the measurement time per class,

f_s : sample rate,

N_{sps} : number of samples per segment, and

N_s : number of segments to be extracted per class.

After segmentation, samples of class are almost identical which is shown in Figure 5.11

For MWF datasets, as only one accelerometer is glued in the work piece and thread forming number for each MWF are the same, so measurement number are the same which are shown in Figure 5.12 and Figure 5.13. Data selection method are the same as in the approach 1. In the step of segmentation, both MWF16 and MWF19 datasets are segmented based on tap speed and its geometrical polygon form. The tap contains five polygons. According to the equation 4.1, the sample rate f_s of 4 MHz and the different rotational speeds in MWF16 (1000 rpm) and MWF19 (1061 rpm), the segment sizes and period duration's for five rounds, one round, and one polygon are shown in Table 4.1. The segment sizes and region duration within the dataset MWF19 and MWF16 are constant, and the number of considered measurements per class is balance. Therefore, the number of segments per class N_s is also balanced if the overlap percentage of all classes is constant. To keep the dataset balanced while multiplying the data, a constant overlap percent over all classes will be used to train and test the most promising classification approaches. The value of the overlap percent depends on the segment size. For non-stationary vibration signals, features in time-frequency domain would be much easier to analyze than just

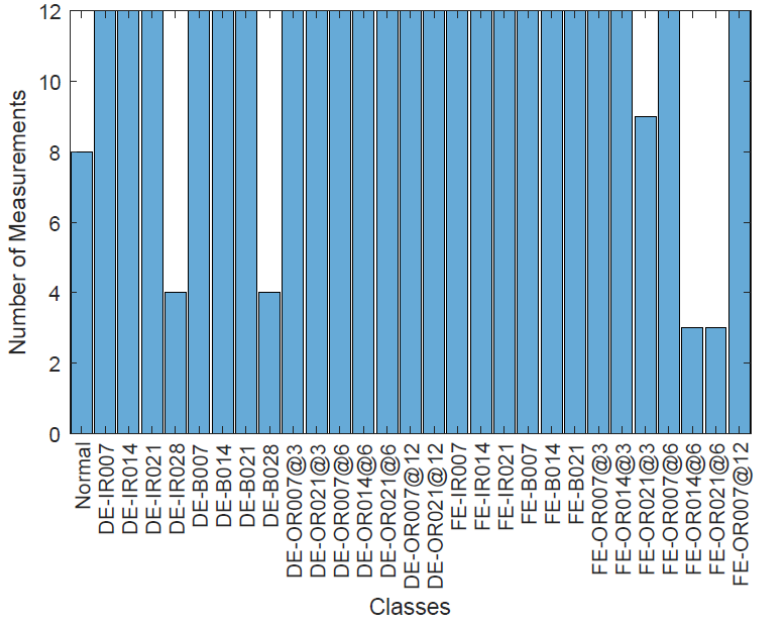

Figure 5.10: Measurements of each class for CWRU bearing dataset [WJS22]

Table 5.1: MWF - segment size and period duration [Joc21]

Segment	Size (samples)		Duration (ms)	
	2016	2019	2016	2019
FiveRounds	1200000	1131008	300	282,752
OneRound	240000	226202	60	56,5505
OnePolygon	48000	45240	12	11,31

analyzing feature in time domain. Therefore, segments are transformed from time domain to time-frequency domain. As introduced in chapter 2, STFT is a suitable way for segment transformation. After segments are transformed by STFT, corresponding spectrograms are obtained. All considered measurements have a sampling rate of 12 kHz for CWRU bearing dataset, considering the Nyquist rate - the nyquist sampling rate is two times the highest frequency of the input signal, the detectable frequency maximum is less than 6 kHz. Examples of spectrograms for CWRU segments are shown in Figure 5.14. For MWF datasets, allmeasurements have a sampling rate of 4M Hz, correspondingly, the detectable frequency maximum is less than 2M Hz. Examples of spectrograms for CWRU segments are shown in Figure 5.15. Parameters in STFT are needed to optimized as they have great impact on spectrograms. The details of parameters are shown in Table 4.2. Data normalization is used to flatten value differences inside same classes caused by different distances of sensory or other circumstances like wear which lead to outlier values in a measurement. Normalization makes the input of a classification model better comparable and can also reduce the computational load [WJS22]. Three options for applying normalization technologies in the proposed approach are known:

i) In the process of segmentation, normalization methods are employed when some

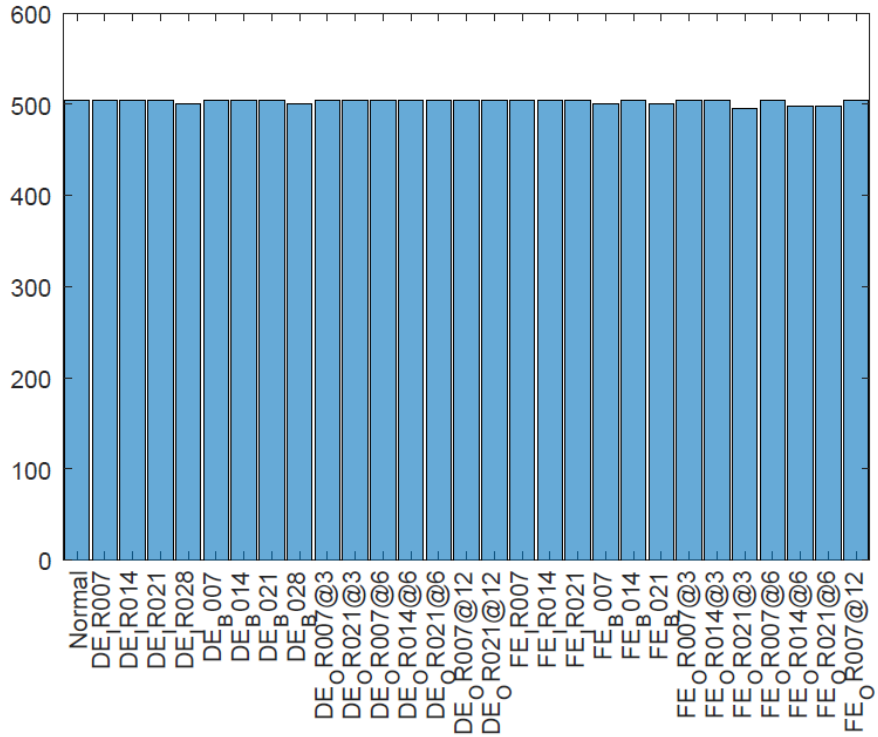

Figure 5.11: Measurements balance of CWRU bearing dataset

segments' value is much larger or smaller than others. While only part of segments are normalized instead of all signals in this step, normalization in this step is denoted as Norm Individual Signal here. Z-Score normalization method is applied in this step. Segments have a mean of 0 and standard deviation of 1 after this normalization step [WJS22].

ii) After measurements are segmented, data in each segment are transformed from time domain to time-frequency domain by STFT. A transformed segment is named as a predictor in this study. Even though segments are normalized in the previous step, corresponding predictors are still vary. Considering differences among predictors, second normalization including Z-Score and Min-Max normalization technology are employed. Data normalization in this step is denoted as Norm Individual Predictor [WJS22].

iii) To reduce the computational load, all predictors can be normalized with Min-Max normalization before predictor features are extracted and classified using CNN. Normalization in this step is denoted as Norm Overall Predictors.

Motivated by known powerful image classification functionalities, CNN is applied to extract features of normalized spectrogram and differentiate them. Hyperparamters in CNN are divided into variables determining a network structure and variables determining a training algorithm. Neural network structure model needs professional design which is time-consuming, labor-intensive, and the efficiency is low. To reduce computation time and improve efficiency, the Basic6 algorithm which is inspired by the speech recognition example [Sch99] is applied in this study. Structure and layers of the Basic6 network are shown in Figure 5.16. Beside the input layer, features are extracted by six convolution (con) layers which are individually followed by batch normalization (norm), ReLU (relu)

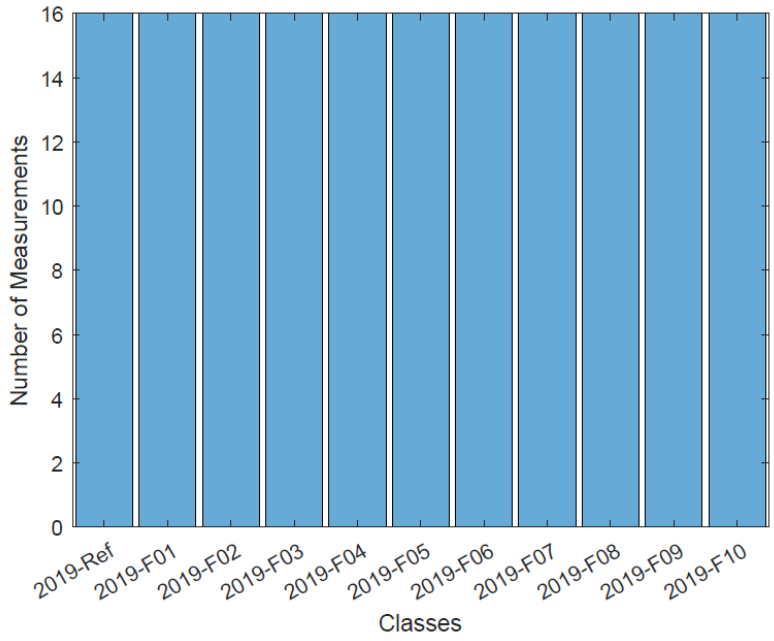

Figure 5.12: Measurements balance of MWF19

activiation function, and max pooling (pool) layers. Three dropout layers are applied between convolutional layers 2 to 5 to prevent the model from over fitting. Classification task is done by a fully connected (fc) layer which has as many neurons as class numbers, followed by a softmax and classification output layer. Hyperparameters related to the training algorithm will be introduced in the next subsection as they can be optimized with parameters together.

Parameters in data processing such as: measurement segmentation, segment transformation, data normalization has a great significance for feature extraction. Meanwhile, hyperparameters in CNN also has great impact on classification results. Parameters in data processing and hyperparameters in CNN can be tuned step by step like in approach 1, however, best combination of them can not be checked when parameters and hyperparameters are optimized separately. In this approach, parameters in data processing and hyperparameters in CNN are optimized in one step as one objective function. Detailed parameters/hyperparameters type are shown in Table 5.2. According to authors information, this is the first time that all parameters and hyperparameters tuning together in one step. In parameters and hyperparameters optimization step, firstly, a set of value with a relatively large span interval is assigned to each Parameter and hyperparameter. Exhaustive sweep optimization are applied to find the good combination of them according to results. In addition, good value interval for a single parameter can also be settled down by the results. Furthermore, Bayesian optimization technology is applied to single parameter and hyperparameter to find best value of them. In this way, best combination of parameters in data processing and hyperparameters in CNN are defined.

Table 5.2: Parameter and hyperparameter option for optimization [WJS22]

No.	Stage	Parameter/hyperparameter type
1	Segmentation	SegmentSize
2		SegmentOverlap
3	STFT	FrequencyLimitMin
4		FrequencyLimitMax
5		PowerScale
6		TimeResolution
7		MinThreshold
8		WindowSize
9		WindowType
10		FrequencyResolution
11		OutputType
12	Normalization	NormIndividualSignal
13		NormIndividualPredictor
14		NormOverAllPredictors
15	CNN	Optimizer
16		MiniBatchSize
17		MaxEpochs
18		ValidationFrequency
19		ValidationPatience
20		InitialLearnRate
21		LearnRateDropPeriod
22		LearnRateDropFactor

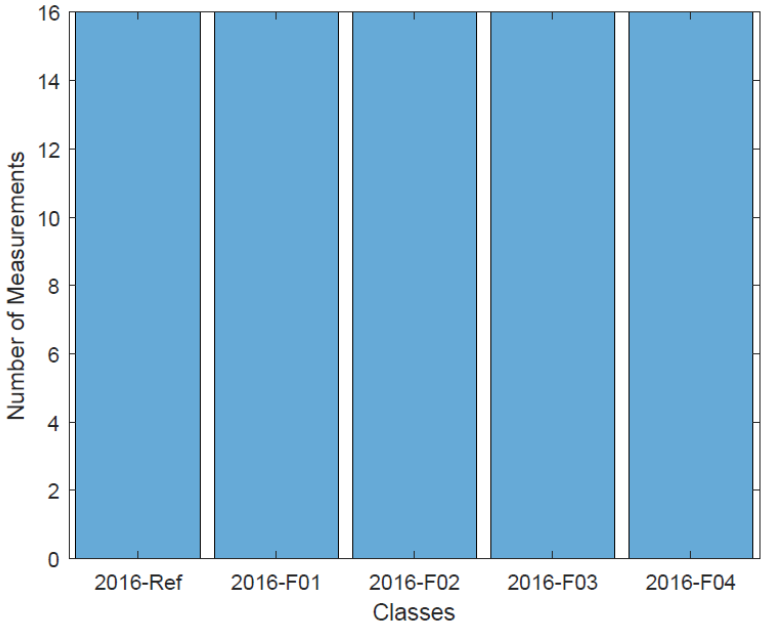

Figure 5.13: Measurements balance of MWF16

5.2.2 Approach 2.0: Transfer learning from MWF19 to MWF16

Steps in approach 2 for all CWRU, MWF19, and MWF16 datasets are the same. Since data in CWRU bearing dataset are very different form MWF19 and MWF16, so most parameters and hyperparameters are also various between CWRU and MWF datasets. However, as data in MWF19 and MWF16 are very similar, parameters and hyperparameters used for MWF19 can be transferred to MWF16 [WJDS22]. The transfer learning approach between MWF19 and MWF16 is named as approach 2.0 as it is based on approach 2. Workflow of approach 2.0 is shown in Figure 5.17. Tool speed is different between these two experiment: tool speed is 1061 rpm in MWF19 while it is 1000 rpm in MWF16. So, data in each segments are not identical in these two datasets. In addition, 11 kinds MWF are applied in experiment 2019 while 5 kinds MWF are used in experiment 2016, therefore, the classes number are different. Except segment length and classes number, other parameters and hyperparameters are the same between these two datasets [WJDS22]. Therefore, based on those optimized parameters and hyperparameters in MWF19, only two parameters need to changed for MWF16: speed and classes number.

5.3 Approach 3: Integrate various data selection, data processing, and machine learning methods

Data are selected from physic point of view in both approach 1 and approach 2, however, whether other data are suitable for data distinguish is not verified. In addition, methods for data transformation and algorithms for data classification are homogeneous in previous two approaches. Unlike approach 1 and 2, different parts data are selected, various data processing in one step, and machine learning methods are experimented in approach 3. Outline of approach is shown in Figure 5.18. Furthermore, according to the destination,

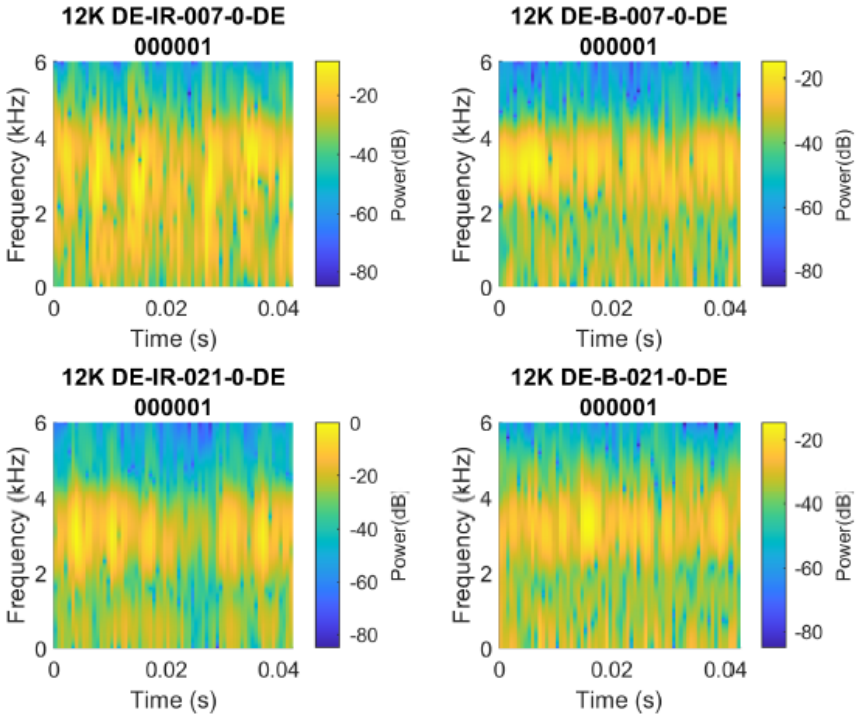

Figure 5.14: Spectrogram of CWRU segments

approach 3 is divided into supervised learning (approach 3.1) and unsupervised learning (approach 3.2). Therefore, instead of a single approach, approach 3 is an integration of various approaches.

Comparing with approaches from other literature, approach 1, and approach 2, highlight of approach 3 is as follows:

i) Data are not selected from physical point of view, various parts data are use to check which part data are more suitable for distinguish.

ii) Various methods are compared in one step in data processing to verify which method is better.

iii) A diversity of ML algorithms are used for feature distinction to search for the best suitable one.

iv) Both supervised learning and unsupervised learning approaches are employed.

5.3.1 Approach 3.1: Supervised learning for IS dataset

Structure health monitoring (SHM) is the original task at the beginning of this project, so CWRU and MWF datasets relevant to SHM are calculated at the beginning. Inner speech dataset was open to public at 2020, it is the last calculated dataset. As inner speech belongs to biology field and it is very new for the author, so this dataset is calculated separately. The approach 3.1 is pinpoint to IS dataset.

In IS dataset, each trial contains data from 128 channel. A few information is relevant to the relations of these 128 channels. Two ways are tried to separate the 128 channels data. The first way separating 128 channels data of each trial into 128 samples and then

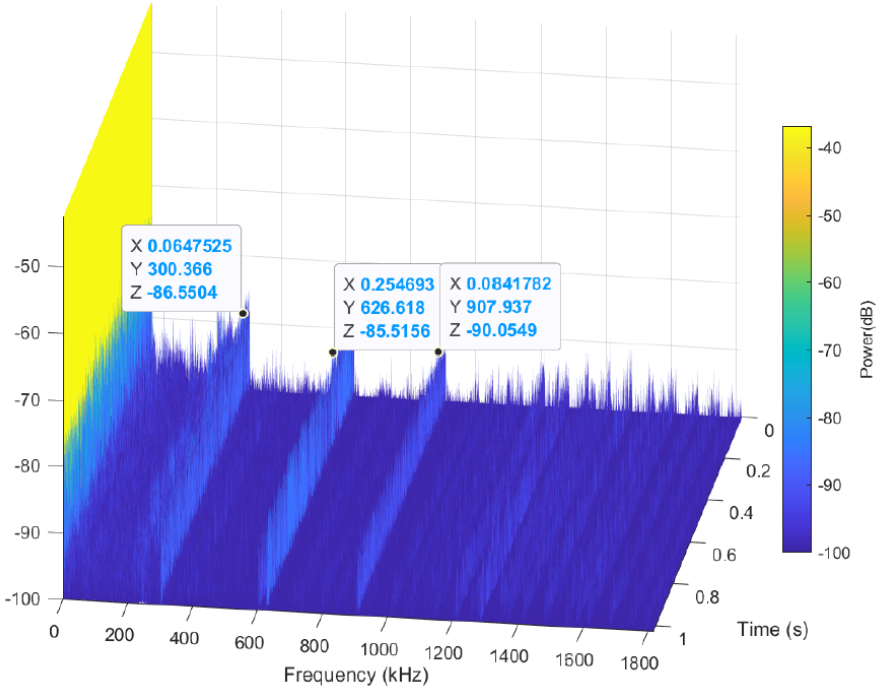

Figure 5.15: Spectrogram of MWF segments [Joc21]

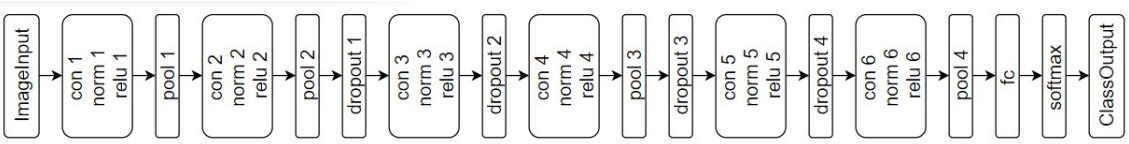

Figure 5.16: Structure of Basic 6 [Sch99]

mix the samples together. Afterwards, the samples are divided into training and test parts. The second way is spliting trails into training-test parts and then separate each trial into 128 channels. Besides, as not all data are referring to IS activity according to literature review. Correspondingly, data should be selected and data from electrodes irrelevant to IS should be dropped. Furthermore, inner speech activity is focus on 1 s - 3.5 s in each trial, whether data in this periods are more suitable for classification is not checked. So in the step of data selection, diverse data are trialed. In addition, both SG filter and EMD method are examined in data processing. Furthermore, various SVM kernels are applied to validate which kernel is more suitable for IS data distinction. This approach is named as approach 3.1 and it's workflow is shown in Figure 5.19.

Data selection

When data from each electrode are used to differentiate the EEG signals, results are poor and scattered as shown in Figure 5.20. The best results for single electrode is less than 32 %. At the same time, it is really time consuming and unnecessary to calculate data form all electrodes as speech related neurons do not cover the whole cranial. Consequently, data selection is necessary for data calculation. According to [CK10], the ability to decode

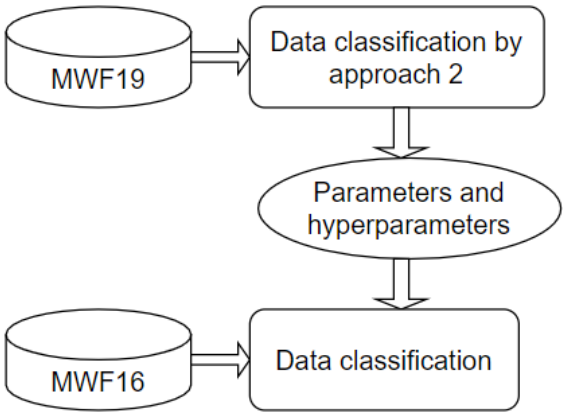

Figure 5.17: Transfer learning from MWF19 to MWF16 [WJDS22]

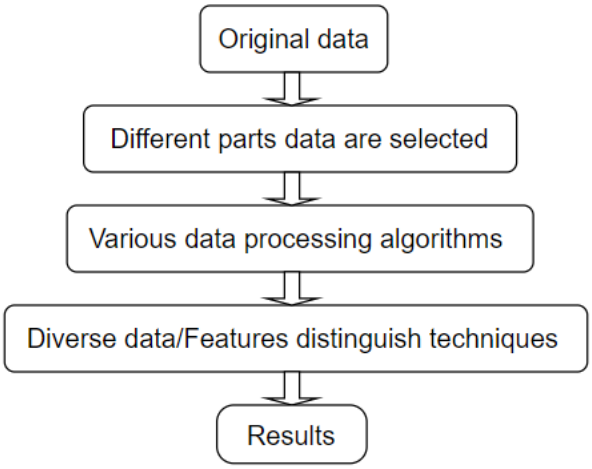

Figure 5.18: Workflow of approach 3

speech in a meaning manner is a complex task that involves multiple stages of neural processing. Around 90 % of humans prefer their right hand for unimanual actions and are left-hemisphere dominant for language functions [PP11]. In the past, it was supposed that language was associated with the activity of three areas in the left hemisphere: the posterior frontal lobe, the upper segment of the temporal lobe, and the insula. Nowadays, most researchers tend that the cortical regions of the brain associated with the comprehension of language are Wernicke's area and the Broca's area [DWVVJ+04]. Wernicke's area controls all language comprehension while Broca's area all language production and that the transmission of information between these areas is facilitated by the arcuate fasciculus [Dro00]. Besides, it is rational to consider that the primary motor cortex may also show similar hemispheric specialization for speech production [TFS+06]. Representation of continuous speech in the primary auditory cortex neurons is measured and how individual phonemes modulate activity across the population of auditory is examined in [MDS07].

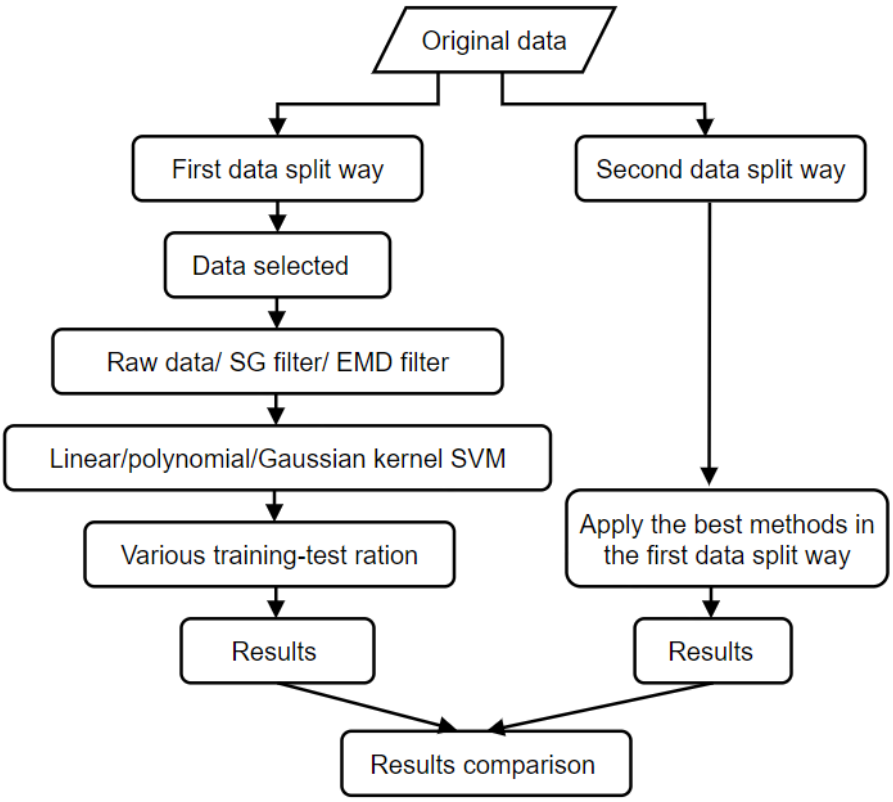

Figure 5.19: Workflow of approach 3.1

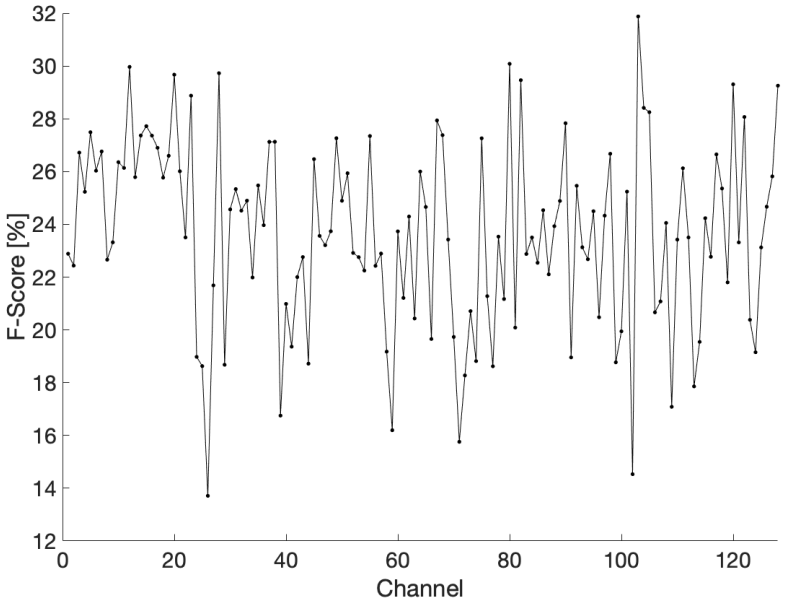

Figure 5.20: Results from each channel in IS dataset [WSS22]

Furthermore, the supramarginal gyrus seems to be involved in phonological and articulatory processing of words, whereas the angular gyrus seems more involved in semantic processing [DCR+92]. Combination of different part in cortex in head that relevant to speech is shown in Figure 5.21 left. Correspondingly, this area is from channel D5 to D32 (electrodes inside of red circle) in headcap which is shown in Figure 5.21 right. Subjects action interval for

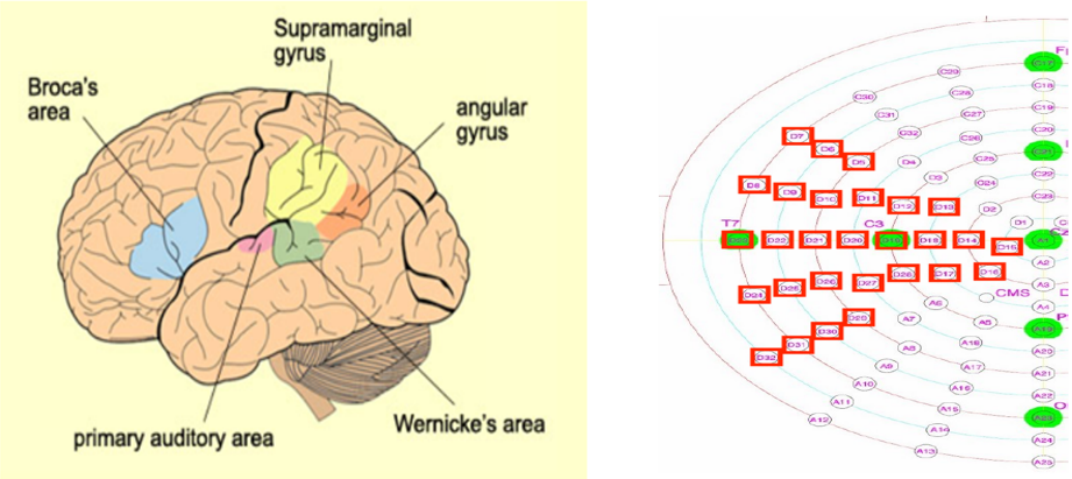

Figure 5.21: Electrodes selection (left: relevant district; right: corresponding electrodes) [Sur22]

IS is only 2.5 s while the first 1 s and the last 1 s are for concentration and relax according to Figure 3.2. Based on the assumption that data activation interval is more involved for IS, data in this part are also applied. Results comparison will verify if activation interval is more suitable for IS distinction.

Data processing

Acharya et al. [ARAS16] conclude that SG filter is a good method for EEG data analysis. Therefore, SG filter is also applied for EEG feature extraction besides to raw measurements. Two variables: the order of polynomial k and the frame length f should be set in the process of SG filter. According to [Sch11], the number of frame length f should be higher than the number of polynomial order $k+1$ to smooth signals by SG filter. Based on these rules, various setting of k and f are experimented. Difference or raw data and after SG filter is presented in Figure 5.22. Global brain activity is conventionally measured using electroencephalograms comprised in oscillations in several functionally-relevant frequency bands [MOM+17]. The bands are identified as δ (0.5-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-35 Hz), γ (35- Hz) waves [GBK+13]. According to [KCJ09], at the end of the arousal spectrum individual is basically disassociated from external world and exhibits a predominance of the δ band. With a predominance of the θ band, the individual focus is on the internal world. A predominance of β waves signal a state encompassing the thinking process with its accompanying ego reactions. The brain waves of α may be considered a bridge from the external world to the internal world. The brain wave band of γ is measured between (35-44 Hz) is the only frequency group found in every part of the brain. Based on the characteristics, selected measurements are decomposed into five IMFs. Raw data

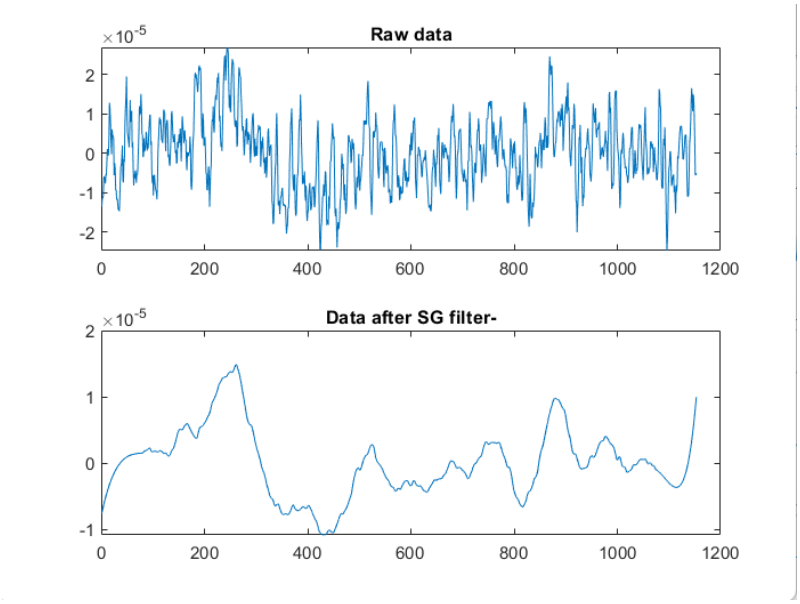

Figure 5.22: Szvitzky-Golay filter application (up: raw data; down: data after SG filter)

and its IMFs are shown in Figure 5.23. Transferring each IMF using Hilbert transform corresponding brain waves bands become visible, they are corresponding to the brain waves bands. That is: γ band results from spectrum of IMF1; β band results from spectrum of IMF2; α band results from spectrum of IMF3; θ band results from spectrum of IMF4; and δ band results from spectrum of IMF5. An example is IMF1 and the corresponding spectrum as shown in Figure 5.24.

Kernels of SVM

Since SVM is a kernel tricky algorithm, kernels have significantly effect on soft margin building. Linear, 2nd, 3rd, 4th polynomial kernels and medium, fine, coarse Gaussian kernels are tried to verify which kernel is more proper for EEG signals in this study.

Training-test ration

To evaluate the performance of a model, observations used in the training process are needed. Otherwise, the evaluation of the model would be biased. The simplest method is to divide the whole dataset into two sets. One is used for training and the other for test (model evaluation). Training set is employed to update model parameters during learning phase and the test set is used to test the ML model after the training phase is complete. Training and test datasets should follow the same distribution. For the IS dataset, various training-test ration - 90:10, 85:15, 80:20, and 70:30 - are trialed to verify models robustness [Sur22].

5.3.2 Approach 3.2: Unsupervised learning for CWRU and MWF datasets

Approach 1 and 2 applied to CWRU bearing dataset and MWF datasets are both supervised learning. In this section, unsupervised learning will be applied to these datasets.

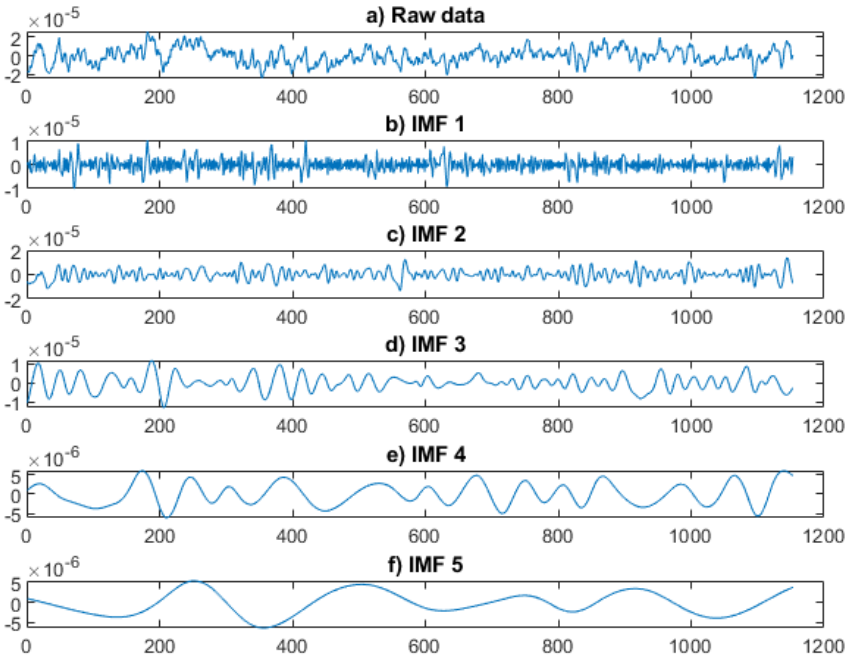

Figure 5.23: Data and corresponding IMFs (a: raw data; b-f: IMFs) [WSS22]

Furthermore, different parts data are selected to verify which data are more suitable for clustering [WLS22]. Moreover, CWT and HHT are also employed for data transform in the step of data transformation. Autoencoder is used as an alternative for features extraction in some trials besides raw segments. Finally, both k-mean and GMM are applied for features clustering. Flowchart of this approach is shown in Figure 5.25.

Data selection

In the approach 1 and approach 2, data are selected based on physical point of view. In approach 3.2, different phases data are compared to check which parts data are more suitable for clustering. Three accelerometers (sensors) are located in basement, nearby driven and fan end in CWRU bearing dataset. Data from one sensor or all data from three sensors are applied in approach 1 and approach 2. However, useful data from nearby sensors or from far away sensors are not verified. Therefore, both data from nearby sensor and far sensors are applied to check which data are more sensitivity. For MWF datasets, in approach 1 and 2, data from forward part are selected for classification from physical point of view. However, whether reverse part data are also available for clustering is still questionable. Both forward and reverse parts data are employed for clustering in this approach. For the bounds among different phases in MWF measurements, the method applied in approach 1 is also used here.

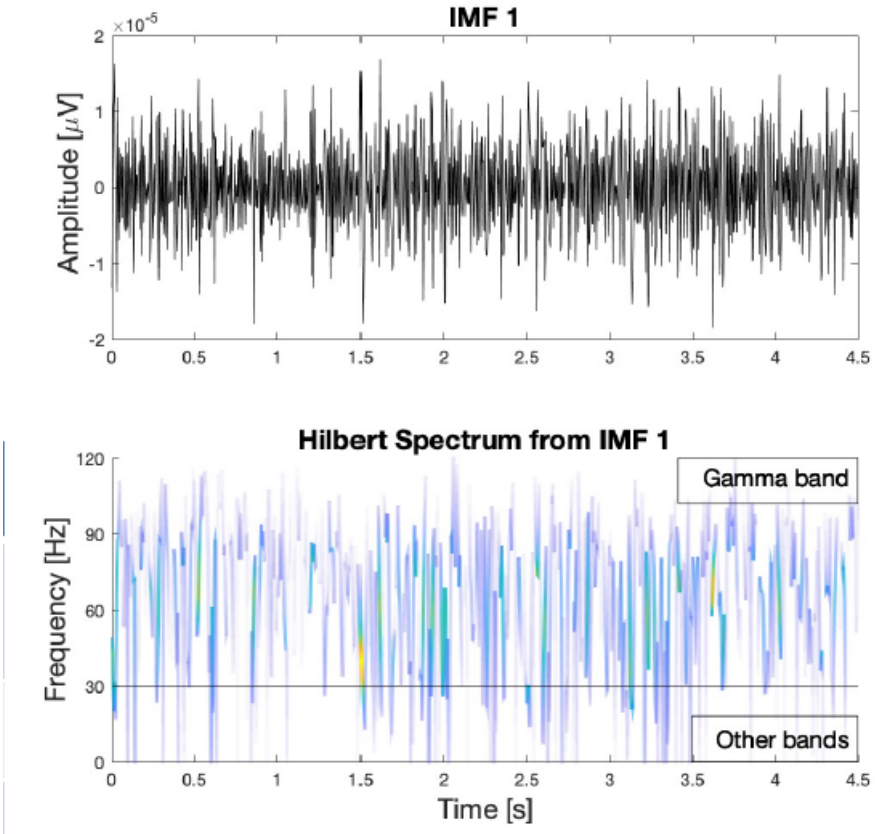

Figure 5.24: IMF and corresponding spectrum [WSS22]

Data processing

For measurements segmentation, one segment contains at least data from one rotation according to tool speed. Unlike segment length is integer multiple of a circle in approach 2, other segment length are trialed in this approach.

Continuous wavelet transform is applied in approach 2 just for searching the boundaries among different parts in time domain [WS22]. However, CWT is also extremely useful in revealing non-linear and non-stationary signals. Since width and height of wavelet are changeable in CWT, as a consequence, time and frequency resolution are very precise. Besides STFT applied in approach 2, CWT is also applied for transforming signals from time domain to time-frequency domain. In approach 3.1, HHT is used for searching brainwaves frequency bands. Hilbert-Huang transform is employed to get spectrum of signals in this approach. Samples of spectrogram and scalogram have been shown in previous section, one sample of spectrum is shown in Figure 5.26. Spectrogram from STFT, scalogram from CWT, and spectrum from HHT are gotten after transformation. These photos can be distinguished directly by clustering algorithms. In case their pixels are very large, to compress the image size and get the main features, autoencoder is applied for data compression as a alternative.

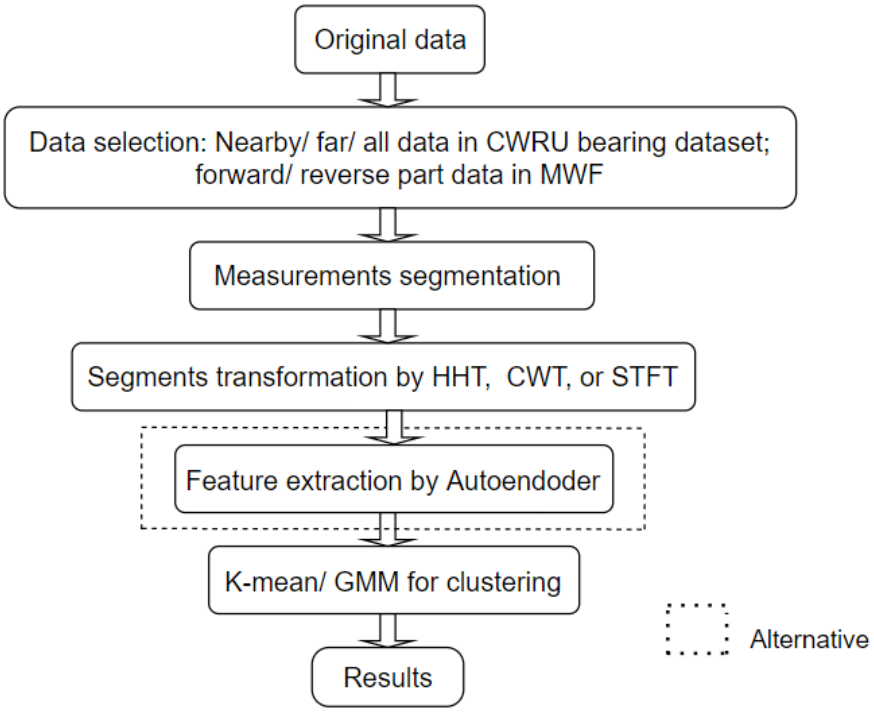

Figure 5.25: Workflow of approach 3.2

Feature clustering

Two kind of clustering methods are applied for feature distinguish. Firstly the classical cluster algorithm K-means are trialed. In addition to K-means, GMM is also applied for features clustering because it has some advantages than K-means. Firstly, Gaussian mixture models do not assume clusters to be any geometry and they work well with non-linear geometric distributions. Besides, GMM has no bias on the cluster sizes. Own to the advantages of GMM, it is also employed in this approach.

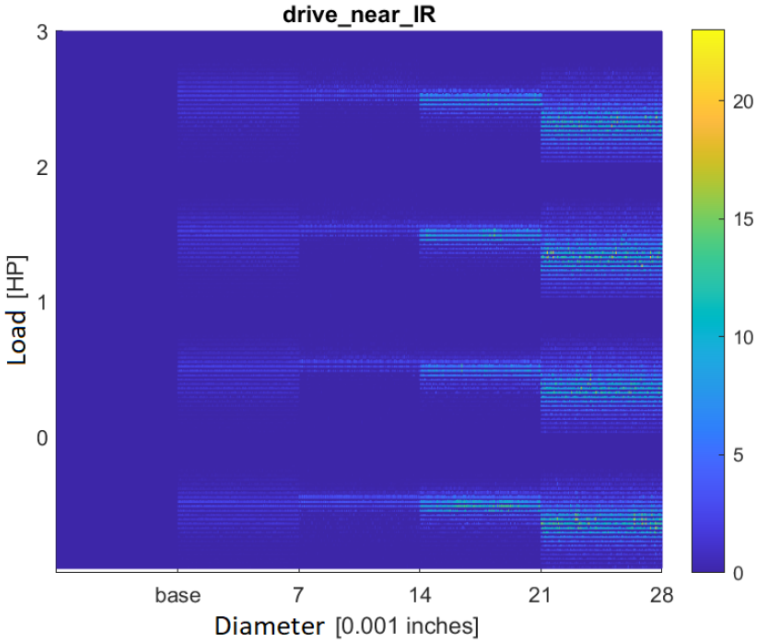

Figure 5.26: Sample of spectrum [Lee21]

6 Results from proposed approaches to relevant datasets

Inner speech dataset is a new dataset as it was open to public at 2020. Since brainwaves are different from vibration and AE signals, supervised learning approach (approach 3.1) is used to distinguish the EEG signals. Comparing with IS dataset, CWRU bearing dataset and MWF datasets have been around for longer years, so both supervised learning approaches (approach 1 and 2) and unsupervised learning approach (approach 3.2) are applied on CWRU and MWF datasets. Furthermore, transfer learning (approach 2.0) is applied between MWF19 and MWF16. As a summary, three approaches (approaches 1, 2, and 3.2) applied on CWRU bearing dataset, three approaches (approaches 1, 2, and 3.2) are applied on MWF19, three approaches (approaches 1, 2.0, and 3.2) applied on MWF16, and one approach (approach 3.1) is applied on IS dataset. Results of these datasets from proposed approaches will be shown in detail.

The contents, figures, and tables presented in this chapter are modified after previous publications [DWS21], [WLS23], and [WS22]. Part of the contents, figures, and tables are prepared for publications of [WSS22], [WLS22], and [WJS22].

6.1 Results for IS dataset

A few publications are relevant to IS dataset as the dataset is very new, therefore, a few information on the way of training-test data split. Two ways of training-test split are tried. The first way is to separate each trial data into 128 channels, mixed the data of each channels together and then divide them into training-test parts. Another way is dividing trials into training-test firstly and then separate each trial into 128 channels.

6.1.1 Results from first data split way

As the headcap contains 128 sensors in the IS dataset experiment, there are 128 channels data for each trial. In the calculation process, data from each channel is treated as one sample firstly, this means that each trial contains 128 samples. Before the trials are separated into training-test parts, each trial is split into 128 samples. Afterwards, the samples from different trials are divided into training and test parts. In this step, different channels data, various data processing methods, varied kernels SVM, and different training-test ratio are tried.

To evaluate the models comprehensively, the results are tested in two schemes: individual scheme and all subjects scheme. In individual model, both training and test data come from the same subject. In all subject scheme, training data come from all subjects and test data come from part of each subject. In addition, both accuracy and F-score are employed to measure trained models comprehensively. Furthermore, to evaluate the deviation of models of each subject, standard deviations is also applied [WSS22].

Results from varied data selection

Linear kernel SVM is the basic SVM, linear SVM is applied for the trials in the beginning. When linear SVM applied to single channel measurements within all subjects scheme, results that shown in Figure 5.20 are not good. After data from channels D5 to D32 are selected, under the training-test data ration with 85-15, results from data within selected 28 channels are shown in Table 6.1. From physical point of view, activity interval of each

Table 6.1: Results from selected electrodes data [WSS22]

Subjects	Individual scheme		All subjects scheme	
	Accuracy	F-score	Accuracy	F-score
Sub-01	91.55	91.55	49.52	49.44
Sub-02	85.52	85.47	51.73	51.73
Sub-03	97.62	97.63	51.83	51.93
Sub-04	97.12	97.09	53.47	53.33
Sub-05	90.67	90.63	50.10	50.05
Sub-06	87.87	87.83	53.93	53.94
Sub-07	93.45	93.42	50.00	49.89
Sub-08	92.86	92.82	52.85	52.65
Sub-09	96.13	96.08	51.20	51.13
Sub-10	98.41	98.40	51.13	51.12
Average	93.12	93.09	51.58	51.52
SD	4.09	4.10	1.42	1.42

measurement should be more relevant to IS. So data in activity part are picked out to verify this assumption. When only data from the activity part (from 1 s to 3.5 s) are applied to for calculation, the results are shown in Table 6.2. Compare with values in Table 6.1 and

Table 6.2: Results from selected electrodes data and activity part [Sur22]

Subjects	Individual scheme		All subjects scheme	
	Accuracy	F-score	Accuracy	F-score
Sub-01	85.48	85.51	43.23	43.23
Sub-02	81.75	81.74	44.03	44.00
Sub-03	93.25	93.20	44.24	44.25
Sub-04	92.36	92.35	45.55	45.47
Sub-05	85.52	85.48	43.74	43.66
Sub-06	81.48	81.43	48.99	48.93
Sub-07	81.25	81.22	44.40	44.37
Sub-08	91.43	91.41	46.67	46.35
Sub-09	92.46	92.37	43.91	43.71
Sub-10	96.92	96.91	46.71	46.59
Average	88.19	88.16	45.15	45.06
SD	5.45	5.45	1.72	1.69

Table 6.2, results from activity part data are not as good as the whole measurements. So the assumption that activity parts data are more relevant to IS is not confirmed.

Results from different data processing

To search for the best features extraction methods, SG filter, and EMD are employed to extract features from raw signals. After features are extracted from these two methods, they are employed as SVM as samples which has been introduced in section 5.3.1. Results for raw signals has been shown in Table 6.1 and 6.2. Results from features after SG filter are shown in Table 6.3. Comparing results in Table 6.1 and 6.3, results from samples after SG filter are not as good as raw signals. Another data processing method is EMD. After

Table 6.3: Results from SG filter [Sur22]

Subjects	Individual scheme		All subjects scheme	
	Accuracy	F-score	Accuracy	F-score
Sub-01	88.58	88.59	46.30	46.19
Sub-02	84.42	84.39	48.17	48.15
Sub-03	97.62	97.65	48.82	48.92
Sub-04	95.83	95.85	50.54	50.39
Sub-05	90.18	90.18	46.34	46.25
Sub-06	88.42	88.42	51.80	51.72
Sub-07	90.58	90.60	46.16	45.93
Sub-08	93.45	93.47	48.22	47.91
Sub-09	95.14	95.13	48.20	48.14
Sub-10	97.82	97.81	48.18	48.15
Average	92.20	92.21	48.27	48.18
SD	4.23	4.10	1.73	1.75

signals are decomposed by EMD, five IMFs are gotten. When these IMFs samples are put into SVM, F-score values from single IMF and combination of IMFs in individual scheme are shown in Table 6.4.

Table 6.4: Results from different IMFs [WSS22]

Subjects	Single IMF				Combination of IMFs		
	1st	2nd	3rd	4th	1st and 2nd	1st to 3rd	1st to 4th
Sub-01	92.02	73.62	50.61	42.61	94.14	94.03	93.09
Sub-02	88.77	65.86	47.15	36.83	88.62	88.99	89.29
Sub-03	97.07	84.82	53.52	45.40	99.09	99.22	99.48
Sub-04	98.82	82.86	58.06	45.26	98.54	98.45	97.73
Sub-05	90.36	74.01	54.77	38.78	91.09	92.37	92.19
Sub-06	89.44	71.88	54.80	40.09	91.76	93.40	92.52
Sub-07	87.88	63.49	45.17	36.20	91.44	93.19	94.40
Sub-08	96.41	82.16	61.43	42.34	96.27	96.88	97.49
Sub-09	95.21	84.64	57.24	45.22	96.28	97.51	96.62
Sub-10	97.40	81.23	57.39	46.13	98.21	99.02	98.63
Average	93.34	76.46	54.01	41.89	94.54	95.13	95.04
SD	3.87	7.41	4.81	3.53	3.47	3.23	3.20

Results from various SVM kernels

All previous calculation are based on the linear kernel SVM. Since SVM is a king of ML that has kernel tricky, different kinds of kernels could be applied. In the step of features classification, various SVM kernels are trialed for the same data. Under the setting of using selected channels data and individual scheme, accuracy value are shown in Table 6.5.

Table 6.5: Results from different SVM kernels [WSS22]

Subjects	Linear kernel	Polynomial kernel			Gaussian kernel		
	Linear	2nd	3rd	4th	Coarse	Medium	Fine
Sub-01	91.55	97.74	98.57	97.86	73.93	98.45	92.90
Sub-02	85.52	92.16	93.25	93.95	64.68	92.26	78.47
Sub-03	97.62	99.34	99.74	100.00	75.40	99.87	98.81
Sub-04	97.12	99.80	99.60	99.60	78.47	99.50	94.51
Sub-05	90.67	97.12	97.82	99.92	78.37	96.03	94.35
Sub-06	87.87	95.59	94.49	94.38	66.92	93.61	66.92
Sub-07	93.45	98.71	99.01	99.31	74.21	98.21	93.25
Sub-08	92.86	98.57	98.69	98.45	74.05	97.98	96.07
Sub-09	96.13	99.80	98.71	99.40	79.66	99.21	97.82
Sub-10	98.41	99.70	99.40	99.11	87.30	98.81	94.44
Average	93.12	97.85	97.93	97.90	75.30	97.39	90.63
SD	4.09	2.29	2.12	2.06	6.08	2.46	9.53

Results from various training-test ration

Whole dataset should be divided into training-test parts: training data are used to train models while test data are applied for testing models. When ration between training and test data are different, classification results are also various. Under the setting: raw whole measurements, 4th polynomial SVM, and all subjects scheme, comparison of F-score from various training-test ration are shown in Table 6.6. The results from various training-test ratio is not significant different.

Best results for the first split way

For IS dataset, data are selected and processed in various ways. In addition, different kernels are applied in features classification. Furthermore, different training-test ratio are also trialed. Therefore, a large number combination of them can be calculated, details results of these combination will not be shown here. From previous calculation, best results come from whole measurements on 28 electrodes, the features that combine IMF1, IME2, and IMF3 in data decomposition step, 4th polynomial kernel in SVM, and 90:10 training-test ratio [WSS22]. Detail value of best results are shown in Table 6.7.

6.1.2 Results from second data split way

The previous calculation steps, firstly the channels of each trial are separated and then data from each channel are mixed together. After that, the whole data are separated into

Table 6.6: Results from different training-test ration [Sur22]

Subjects	Training-test ration			
	90:10	85:15	80:20	70:30
Sub-01	96.63	96.86	96.64	95.62
Sub-02	99.42	99.51	98.67	97.83
Sub-03	99.19	99.33	99.30	98.31
Sub-04	99.70	99.32	98.96	97.75
Sub-05	99.16	99.04	98.47	97.54
Sub-06	99.48	99.32	98.42	98.74
Sub-07	99.21	99.16	98.02	97.74
Sub-08	99.44	99.29	99.30	98.69
Sub-09	99.53	99.51	98.51	97.38
Sub-10	99.40	99.21	98.89	97.84
Average	99.12	99.06	98.75	97.75

Table 6.7: Best results for IS dataset [WSS22]

Subjects	Individual scheme		All subject scheme	
	Accuracy	F-score	Accuracy	F-score
Sub-01	99.64	99.63	99.11	99.09
Sub-02	96.83	96.83	99.85	99.86
Sub-03	100.00	100.00	99.81	99.80
Sub-04	100.00	100.00	99.71	99.71
Sub-05	98.61	98.61	99.58	99.57
Sub-06	98.57	98.57	99.83	99.83
Sub-07	99.90	99.91	99.53	99.54
Sub-08	99.29	99.27	100.00	100.00
Sub-09	99.80	99.81	99.85	99.84
Sub-10	99.80	99.80	99.40	99.40
Average	99.24	99.24	99.67	99.66
SD	0.95	0.95	0.27	0.27

training and test parts. Another way to split data is firstly split trials into training and test parts and then channels data of each trial are separated. When the raw data and data after EMD are split in this way, the results of each subject and results of all subjects from 4 polynomial SVM are shown in Table 6.8.

Table 6.8: Results of second data split way

Subjects	Raw data		IMF1-3	
	Accuracy	F-score	Accuracy	F-score
Sub-01	38.26	32.74	20.12	20.23
Sub-02	19.64	19.67	21.92	21.62
Sub-03	13.23	12.76	16.53	16.43
Sub-04	16.96	17.13	18.25	18.00
Sub-05	25.60	24.57	24.01	23.84
Sub-06	28.13	27.77	22.21	21.99
Sub-07	18.25	18.32	18.85	18.57
Sub-08	27.62	27.19	16.90	17.08
Sub-09	34.13	34.48	39.38	39.41
Sub-10	27.38	27.39	25.20	23.64
Sub 1-10	26.25	26.25	25.65	25.63

The results of each subject for raw data are differ strongly: The accuracy is from 13.23 % to 38.26 % and average F-score is from 12.76 % to 34.48 %. When the raw data are used for classification, the accuracy of all 10 subjects is 26.25 % and the average F-score is 26.25 %. After the raw data are processed by the EMD, the accuracy is from 17.08 % to 39.41 % and the average F-score is from 16.53 % to 39.38 % for each subject. The accuracy of all 10 subjects is 25.65 % and the average F-score is 25.63 %. Comparing the values in different columns, when the whole dataset is split in the second way, there is no improvement comparing with the results from data after EMD and results from raw data. Furthermore, compared with the individual scheme results which also use the raw data in Table 6.1 and the results of EMD data in Table 6.8, the results differ significantly. Therefore, the dataset split way effect the results greatly.

6.2 Results for CWRU dataset

Both supervised and unsupervised learning are applied to CWRU bearing dataset. To compare their results, results from supervised learning are put together and results from unsupervised learning are put together. Detail results are shown in the following sections.

6.2.1 Results from supervised learning

Most approaches applied on CWRU bearing dataset choose the baseline and fault states from driven end and divide these data into 10 classes. While all these ten classes contains data from three channels. To compare with results from other contribution, these three channels data are also employed to distinguish these ten states in approach 1. Besides applying data from three channels, data just from DE channel are also applied to distinct these states. For all 29 states in CWRU bearing dataset, as only DE channel data are

included in all classes, so the DE data are also used for all states differentiation. To reduce the randomness of the results from a single run, data are calculated 5 times and for each time the training data and test data are randomly selected. Detail values of each time and average values of these five times are shown in Table 6.9. Besides accuracy is applied to

Table 6.9: Results from approach 1 for CWRU dataset

Classes	Data used	Channels	Test accuracy (%)					
			1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	average
10	DE, normal	3 channels	100	100	100	100	100	100
10	DE, normal	DE	99.82	99.87	99.82	99.87	99.84	99.84
29	all	DE	94.62	94.48	94.88	93.52	93.37	94.17

evaluate trained models, F-score is also applied as metrics in approach 2. Furthermore, 5-fold cross validation are employed to check the trained models robustness. As accuracy value in approach 1 for 10 classes achieve to 100%, it is no need to differentiate these ten classes again by approach 2. All twenty-nine states in CWRU bearing states are classified by approach 2 [WJS22]. Detail results are shown in Table 6.10. From Table 6.9 and 6.10,

Table 6.10: Results from approach 2 for CWRU dataset [WJS22]

Classes	Data used	F-score (%)						Accuracy
		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	average	
29	all	100	100	100	100	100	100	100

the following conclusions can be drawn:

i) Results from each time calculation have subtle difference no matter on approach 1 or approach 2, this denotes the robustness of trained models.

ii) When ten classes are differentiated, results from approach 1 arrived at 100 % when using data from three channels while 99.84 % when using data from DE channel. These results are better than most other contributions, this denotes that approach 1 outperform than most other contributions. Besides, from results comparison inside approach 1, data from three channels are easier to distinguish than data from one channel.

iii) No matter F-score and accuracy values are 100 % when twenty-nine bearing states are distinguished from approach 2. This means that approach 2 can distinguish all bearing states perfectly. Comparing with results in Table 4.2, approach 2 is the best approach among all contributions.

6.2.2 Results from unsupervised learning

Before samples are put into machine learning, data are analyzed preliminary in unsupervised learning approach 3.2. Signals in time domain are transformed to spectrogram by STFT, scalogram by CWT, and spectrum by HHT. From comparison, near measurements spectrogram and scalogram are more distinguishable than remote measurements. The amplitude range for the near measurements are larger and wider than remote measurements. This phenomena is observed across all TFAs which matches view from physical point (the further a signal travels, the weaker it becomes). Therefore, measurements from near accelerometer are used for bearing states clustering.

Three TFA methods are employed in data transformation step: STFT, CWT, and HHT and correspondingly spectrogram, scalogram, and spectrum are acquired. Compare with other photos, signals time-frequency information are more distinguishable by spectrogram when they are analyzed in the preliminary step. Continuous wavelet transform exhibits higher resolution at the lower frequency rang which resulting bright line near the bottom of the images. Spectrum after HHT exhibits similar frequency resolution compare with spectrogram, but the line are not as well defined as spectrogram. When different combination of TFAs and machine learning methods are applied to differentiate faulty-free and faulty states, the good results are shown in Table 6.11. Aside from two clusters (fault-free and

Table 6.11: Two clusters results from approach 3.2 for CWRU dataset [Lee21]

TFAs	Cluster methods	No.	Results					
			Purity	RI	ARI	NMI	F-m	Average
STFT	K-mean	2	1	1	1	1	1	1
CWT	K-mean	2	0.993	0.994	0.986	0.963	0.995	0.986
HHT	K-mean	2	0.443	0.613	0.000	0.051	0.730	0.362
STFT	GMM	2	0.678	0.762	0.400	0.432	0.845	0.623
CWT	GMM	2	0.995	0.984	0.964	0.923	0.988	0.971
HHT	GMM	2	0.443	0.613	0.000	0.036	0.754	0.369

faulty) distinction, more clusters are also tried. As results come from many combination of TFAs and ML algorithms, it is unnecessary to show all results. Only methods that get good results are presented here. Best results for three, four, and five clusters are shown in Table 6.12. From Table 6.11 and 6.12, the following conclusions can be drawn:

Table 6.12: More clusters results from approach 3.2 for CWRU dataset [Lee21]

TFAs	Cluster methods	No.	Results					
			Purity	RI	ARI	NMI	F-m	Average
CWT	GMM	3	0.900	0.842	0.656	0.690	0.779	0.773
CWT	K-mean	4	0.867	0.643	0.333	0.554	0.578	0.595
STFT	GMM	5	0.673	0.779	0.418	0.594	0.560	0.605

i) When data are divided into two cluster, all result values from method of STFT+K-means are 1. This means that combination of STFT and K-mean methods can distinguish fault-free and faulty bearings totally [WLS23].

ii) When data are differentiated into two, three, four and five groups, results from HHT combining with ML algorithms are not good as CWT and STFT. From this point, HHT is not suitable for features extraction in CWRU bearing dataset.

iii) Result from two clusters are higher than results from more clusters. This imply that faulty and fault-free states are more easier to differentiated than more bearing states.

6.3 Results for MWF datasets

To distinguish AE data in two MWF datasets, both supervised learning and unsupervised learning approaches are employed on them. Supervised learning approaches 1 and 2 are applied on MWF19 while supervised learning approaches 1 and 2.0 are applied on MWF16

for data classification. Unsupervised learning approach 3.2 is used for data clustering in both datasets.

When supervised learning approach 1 is applied to MWF datasets, some basic calculation are conducted to avoid that training and test segments are mixed before the final results are gotten. Firstly, various measurements are employed. When various measurements are employed for calculation, the test data are also varied. For example, in the process of fluid 2 and fluid 3 classification process, different threading holes in each series are employed. The results from various measurements are shown in Table 6.13. Test

Table 6.13: Results of fluid 2 and fluid 3 when different measurements are chosen

Measurements	Test accuracy (%)			
	1^{st}	2^{nd}	3^{rd}	average
8-12	84.36	87.65	87.65	86.55
4-12	84.07	82.06	80.65	82.26
4-16	80.05	80.95	84.22	81.74

accuracy from measurements 8-12 are from 84.36 % to 87.65 %, the difference among the results in each time is not great. The average accuracy from measurements 8-12 is 86.55 %, the average accuracy is 82.26 % when measurements 4-12 are selected, and the average accuracy is 81.74 % when measurements 4-16 are chosen. The results are not significant different when various measurements of each series are choose. Although the results are slightly higher than other two ways when measurements 8-12 are chosen, the reason can be explained from pyhsical point of view. As measurements 8-12 are not close or far to the sensor, so less noise are included in these measurements compare with other measurements.

Secondly, to verify the influence of training:test segments partition methods on results, segments are split in two ways: firstly, training and test segments are choose randomly, 70 % are employed for training, 15 % are employed for validation, and the rest 15 % are employed for test. The second way is: the whole dataset is calculated by 10-fold cross validation. Then, the results from training:test data are choose randomly are compared with results from K-fold cross validation. For example, when fluid 2 and fluid 3 in MWF19 are classified, the results from training:test randomly are shown in Table 6.14 (to verify if results are stable, data are calculated 5 times, each time the data are shuffled). The test accuracy are from 95.59 % to 100 %, there are no significant difference among each time calculation. When the whole data are split into ten folds, nine folds are employed as

Table 6.14: Results of fluid 2 and fluid 3 when training-test data split randomly

MWF	Test accuracy (%)					
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	average
f2/f3	98.53	98.53	95.59	95.59	100	97.65

training and the rest one fold segments are applied as test segments. The test and training segments are totally differentiated. The test results of each fold are shown in Table 6.15. When data are split into training:test randomly, the results are from 95.59 % to 100 % and the average is 97.65 %. Results difference of each time is not significant. When data are split into 10-fold cross validation, the accuracy is from 91.11 % to 97.83 % and average is 95.61 %. Small difference are among results of each fold. No matter how segments are

Table 6.15: Results of fluid 2 and fluid 3 when training:test data split by 10-fold cross validation

Test accuracy (%)										
1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	average
91.11	97.83	97.83	93.48	95.65	93.48	97.83	97.78	93.33	97.78	95.61

split, no significant difference are among results from these two ways. This means that when various segments are applied for test, the results are not influenced significantly.

6.3.1 Results for MWF19

Results from supervised learning approaches

Eleven kinds of MWF are employed in experiments MWF19. When approach 1 is applied to MWF19 dataset, these 11 MWF are classified into two ways. The first way is divided 11 MWF into 11 classes - each class stands for one type MWF. In the second way, these 11 types of MWF are divided into 6 big categories according to their additives similarity [DWS22]. Based on Table 3.4, fluid 2 and 3 belong to one group; fluid 4, 5, and 6 belong to one group; fluid 7 and 8 belong to one group; fluid 9 and 10 belong to one group; reference and fluid 1 belong to two groups separately. Then each big category is further classified [DWS22]. To verify the models' robustness each classification step is calculated 5 times and the average value is also calculated. Detail results are shown in Table 6.16. When

Table 6.16: Results for MWF19 dataset from approach 1 [DWS22]

Classes	MWF	Test accuracy (%)					
		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	average
11	Ref./f1/f2/f3/f4/ f5/f6/f7/f8/f9/f10	95.77	93.65	97.88	97.88	97.88	96.61
6	Ref./f1/f2,3/f4,5,6/f7,8/f9,10	96.44	97.33	96.44	97.78	96.00	96.80
2	f2/f3	98.53	98.53	95.59	95.59	100	97.65
3	f4/f5/f6	96.15	94.23	98.08	94.23	93.27	95.19
2	f7/f8	100	100	100	100	100	100
2	f9/f10	98.59	98.59	100	100	100	99.44

approach 2 is employed to the dataset, 11 MWF is divided into 11 classes directly to reduce the calculation time. Besides accuracy, F-score is also applied to evaluate trained models. Detail results are shown in Table 6.17. Based on the results from approach 1 and approach

Table 6.17: Results for MWF19 dataset from approach 2 [WJDS22]

Classes	MWF	F-score (%)					Accuracy (%)
		1^{st}	2^{nd}	3^{rd}	4^{th}	average	
11	Ref./f1/f2/f3/f4/ f5/f6/f7/f8/f9/f10	98.58	98.80	98.15	98.92	98.61	98.58

2, the following conclusion can be drawn:

i) All results values are over 95 % no matter from approach 1 and approach 2. This imply that both approaches can classify AE signals in MWF19 dataset well.

ii) No big difference among values from each calculating time, this entail the robustness of trained models.

iii) When approach 1 is applied on MWF19, no matter these eleven MWF are divided in the first way or second way, the average accuracy are high. This verify the robustness of this approach.

iv) When fluid 7 and fluid 8 are distinguished by approach 1, all results value are 100 %. This indicates that approach 1 can differentiate fluid 7 and fluid 8 perfectly.

v) Compare approach 2 and approach 1, approach 2 can distinguish these eleven MWF better when all MWF are divided into 11 classes.

Results from unsupervised learning approach

Since diversity data, various segments length, varied data transform methods, and different feature extraction algorithms are applied in approach 3.2. Combination of these ingredients lead to a large number of results. It is impossible to present all results here. Only some good and typical results with specified settings are shown here. Under these setting: K-mean as cluster, segment length is 1000000, STFT is settled as TFA, good results are shown in Table 6.18. Based on the calculation process and results in Table 6.18, the following

Table 6.18: Results for MWF19 dataset from approach 3.2 [Lee21]

Phase	Encoded	Clustering mode	Test average
reverse	no	ref *vs* fluid 1	0.29
reverse	no	ref *vs* fluid 2 - 10	1.00
last second of forward part	yes	ref *vs* other fluids	1.00
forward	no	ref *vs* fluid 4	1.00
forward	no	ref *vs* fluid 5	0.94
forward	no	ref *vs* fluid 9	0.94
forward	no	ref *vs* fluid 6	0.91

conclusion can be drawn:

i) Best results come from given setting for this dataset - segment length is 1000000, K-means as cluster, STFT as TFAs.

ii) When reverse part data are analyzed, all metrics values are 1.0 from reference fluid and fluid 2-10. This suggest that the approach 3.2 can distinguish reference fluid and fluid 2-10 totally [WLS22].

iii) When spectrogram are applied directly instead of compressed by encoded, results are better. This imply that feature compression is not suitable for data extraction in this dataset.

At present, MWF distinction using AE signals is not a common topic, only a few papers focus on differentiating AE signals from various MWF. Wirtz et al. [WDS17] analyze the AE signals from two kinds of MWF using continuous wavelet transform (CWT) and K-means. The accuracy value ranges from 75 % to 87 %. Demmerling et al. [DWS22] distinguish the AE signals from two or three kinds of MWF applying CWT and K-means. The accuracy values is about 40-88 %. Compared with these two contributions, the advantages of the approach 3.2 are:

i) More metrics are applied in this study. Five different metrics are employed to evaluate the approach in this study while there are only cluster photos in other contributions [WLS22].

ii) Reference fluid and fluids 2-10 are distinguished perfectly.

All metrics values from reference fluid to fluid 2-10 clustering are perfect. This indicates that AE signals from reference fluid and fluid 2-10 can be differentiated totally. One exception comes from clustering results of reference fluid and fluid 1. Metrics values from reference fluid and fluid 1 clustering are not ideal. For the reasons of these non-ideal results, the following assumptions from physical point of view are supposed.

i) Measurements locations that filled with reference fluid and fluid 1 are close with each other. Therefore, in the process of thread forming in series m01, the micro changes in these holes affect AE signals in series m02 or vise versa. In other words, AE signals from series m01 and m02 are influenced mutually, so they are not easy to distinguish [WLS22].

ii) Distance from measurements in series m01 to sensor and distance from measurements in series m02 to sensor are very similar. A possibility is that in the process of acquiring AE signals, sensor is not so sensitive that it measures AE events in both series [WLS22].

iii) Additives in fluid 1 (sodium sulfonate) has tiny influence comparing with reference fluid in thread forming process. Effect of additives in fluid 1 are not as strong as additives in fluids 2-10 for the AE signals. Therefore, it is not easy to detect the AE signals difference [WLS22].

6.3.2 Results for MWF16

Results from supervised learning approaches

Similar with MWF19 calculating process, five type of MWF are categorized into two ways when approach 1 is applied to MWF16 dataset. Firstly, 5 MWF is categorized into 5 classes directly. The second way is to divided these 5 MWF into 3 big groups: reference, oil-based, emulsion-based MWF. Then oil-based MWF is sub divided into oil 1 and oil 2; emulsion-based MWF is sub divided into emulsion 1 and emulsion 2 [WS22]. As the same way for calculating MWF19, to verify the models' robustness each classification step is calculated 5 times and the average value is also calculated. Detail results are shown in Table 6.19. Unlike approach 2 which are designed to MWF19 directly, transfer learning approach

Table 6.19: Results for MWF16 dataset from approach 1 [WS22]

Classes	MWF	Test accuracy (%)					
		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	average
5	Ref./emul.1/emul.2 oil 1/oil 2	97.70	98.28	98.85	99.43	96.65	98.11
3	Ref./emul.1,2/oil 1,2	99.04	99.04	99.04	99.04	98.56	98.94
2	emul.1 /emul.2	98.56	96.40	97.84	94.96	100	97.55
2	oil 1/ oil 2	99.28	98.65	97.12	97.84	98.56	98.29

(approach 2.0) is applied to MWF16. It denotes that parameters in data processing and hyperparameters in CNN are not optimized according to this dataset, all parameters and hyperparameters are from approach 2 which are trained from MWF19. Therefore, results for this dataset is not as high as MWF19. Detail value are shown in Table 6.20. The

Table 6.20: Results for MWF16 dataset from approach 2.0 [WJDS22]

Classes	MWF	F-score (%)					Accuracy (%)
		1^{st}	2^{nd}	3^{rd}	4^{th}	average	
5	Ref./emul.1/emul.2 oil 1/oil 2	70.30	91.30	94.51	91.30	86.85	86.20

following conclusion can be drawn when combining results from Table 6.19 and Table 6.20:

i) All results value are over 97 % when approach 1 is applied to this dataset. This imply that approach 1 is suitable for both MWF19 and MWF16. Metalworking fluids are distinguished well using approach 1.

ii) No matter five MWF are classified in the first way or second way from approach 1, the average accuracy are high. This verify the robustness of approach 1.

iii) Comparing results value from approach 2, results from transfer learning are not good as approach that specific to one dataset. This denotes that parameters and hyperparameters optimization for one dataset has significantly influence.

Results from unsupervised learning approach

As explained in the previous subsection, while a large number of combination in each step, it is impossible to present all results from approach 3.2. Only relatively good and some typical results with specified settings are shown here. Under these setting: K-mean as cluster, segment length is 226200, autoencoder as feature extraction algorithm, part of relatively good and typical results are shown in Table 6.21. The following conclusion can

Table 6.21: Part results for MWF16 dataset from approach 3.2 [Lee21]

Phase	TFAs	Clustering mode	Test average
reverse	CWT	ref *vs* emul. 1	0.62
reverse	CWT	ref *vs* emul. 2	0.67
reverse	CWT	ref *vs* oil 1	0.65
reverse	CWT	ref *vs* oil 2	0.61
forward	CWT	ref *vs* emul.1	0.55
forward	CWT	ref *vs* emul.2	0.48
forward	CWT	ref *vs* oil 1	0.42
forward	CWT	ref *vs* oil 2	0.47
reverse	STFT	ref *vs* emul. 1	0.38
reverse	STFT	ref *vs* emul. 2	0.37
reverse	STFT	ref *vs* oil 1	0.43
reverse	STFT	ref *vs* oil 2	0.36
forward	STFT	ref *vs* emul. 1	0.41
forward	STFT	ref *vs* emul. 2	0.42
forward	STFT	ref *vs* oil 1	0.31
forward	STFT	ref *vs* oil 2	0.32

be drawn from Table 6.21:

i) Relatively good results come from given setting for this dataset - segment length is 226200, K-means as cluster, encoder as feature extraction.

ii) Continuous wavelet transform is more suitable for feature extraction than STFT and HHT when specify to MWF16 dataset.

iii) Data from reverse part are easier to cluster than data from forward part.

7 Summary, conclusions, and outlook

7.1 Summary

As machine learning is able to improve calculation continuously, fulfill various decision-making tasks automatically, and identify trends and patterns, it has been widely applied in various fields. Many new designed and pre-trained models are available nowadays. The main problem for most ML approaches is that they are specify to one dataset, their generalization can not be verified. Based on the stat-of-art ML approaches, several improved ML approaches are proposed and their performance are verified by datasets from different fields.

Before proposed approaches are presented, theoretical background relevant to this study is introduced briefly. Data processing methods like data selection, SG filter, segmentation, STFT, CWT, EMD, HHT, and data normalization technology are presented. In addition to data processing methods, ML algorithms like CNN, SVM, autoencoder, K-means, GMM, and TL are introduced. Furthermore, parameters and hyperparameters optimization technology and metrics evaluating machine learning approaches are also presented.

Four datasets from various fields - EEG signals from IS dataset, vibration signals from CWRU bearing dataset, and AE signals from two MWF datasets - are used for validating proposed approaches. Related test rig and experiments procedure are varied greatly. So these datasets and their experiments are introduced in chapter 3. Before approaches are proposed, state-of-art approaches applied in these four datasets are reviewed and their results are exhibited. In addition, problems and disadvantages of these state-of-art approaches are uncovered.

Based on the drawbacks of approaches in literature review, three big categories approaches are proposed. These three categories approaches are further divided into five sub approaches. Flowchart, highlight, and innovation of these five approaches are explained in detail.

i) The approach 1 is mainly focus on CNN hyperparameters optimization. Less data processing technology are used. A new data processing method is trialed - CWT is used to find boundaries among different phase in MWF signals. Besides, measurements are partitioned according to tool speed. Convolutional neural network structure and hyperparameters referring to training algorithms are optimized specific to each dataset.

ii) More data processing methods are applied in approach 2 comparing with approach 1. Data selection method is the same as in approach 1. Segments length and overlap among segments are not fixed in this approach. Segments length is one parameter that can be optimized together with other parameters in data processing. Furthermore, segments are transformed from time domain to time-frequency domain by STFT and spectrograms are acquired. Data normalization methods are applied to segments and spectrograms to remove the outliers. Additionally, unlike hyperparameters are tuned manually in approach

1, parameters and hyperparameters are optimized automatically in one step in approach 2 and a loop among each step is build. To evaluate the approach performance more comprehensively, besides accuracy, F-score is applied as metrics. Based on the approach 2, as MWF 19 and MWF16 are similar with each other, a transfer learning approach (approach 2.0) is raised between two MWF datasets.

iii) Only one method is applied in data processing and distinguishing in each step in approach 1 and 2. On the contrary, several methods are trialed in each step in approach 3. Based on the different destination, approach 3 is sub divided into approaches 3.1 and 3.2. Approach 3.1 belongs to supervised learning while approach 3.2 is unsupervised learning. In approach 3.1, Firstly, training:test data are split into different ways. Besides, whole measurement and activity part of each measurement data are trialed to find the most suitable data. Raw measurement, SG filter, and EMD are employed on data processing to check which is more useful for feature extraction. Various kernels such as linear, Gaussian, and polynomial kernels are also trailed. Besides, various training-test ration are also trialed in approach 3.1. For approach 3.2, near and remote data in CWRU bearing dataset, forward and reverse phase data in MWF datasets are picked out separately in data selection step. Diverse transform technology like STFT, CWT, and HHT are used for acquiring signals frequency bands information. Besides, autoencoder is employed for features extraction as an alternative. Lastly, K-means and GMM are employed to cluster features in approach 3.2.

Five sub approaches are designed and they are validated by four datasets. Consequently, a large number of results are obtained. To state results unambiguously, results are shown based on each single dataset. As IS dataset is new and it is open to public in last two years, so approach 3.1 is applied to it. The results of IS dataset has significant difference when training:test data are split in different ways. For CWRU bearing dataset, the best classification results is 100 % for twenty nine classes by approach 2. All clustering results are 1.0 when faulty and fault-free bearings are clustered. For MWF19 dataset, best classification results are F-score is 98.61 % and accuracy is 98.58 % from approach 2. When approach 3.2 is applied to distinguish features in MWF19, reference fluid and test fluids 2-10 are differentiated totally. For dataset MWF16, the best classification results are accuracy is 98.11 % from approach 1. Although results from approach3.2 for MWF16 is not ideal, some conclusions can also be drawn from calculation process. In addition to results from proposed approaches are presented, results from other literature are also reviewed. By results comparison with other literature, advantages of proposed approaches are obvious.

7.2 Conclusions

Some conclusions can be drawn from single dataset and some general conclusions can be drawn from these four datasets. Conclusions for each database and conclusions from all datasets are presented here.

According to the results comparison with two data split way for IS dataset, when data are splitted in the first way, results are much better than the second way.

From results and results comparison with other literature on CWRU bearing dataset, such conclusions could be drawn: First of all, although approach 1 performs not as good as approach 2 when differentiating all bearing states, it outperform than most other approach when ten bearing states are classified. Then, from results comparison inside approach 1,

data from three channels are easier to distinguish than data from one channel. In addition, approach 2 distinguish all bearing states perfectly. It is the best approach applied to CWRU bearing dataset. Furthermore, when unsupervised learning is applied to CWRU bearing dataset, combination of STFT and K-means can distinguish fault-free and faulty bearings totally. This denotes that comparing with other unsupervised learning, approach 3.2 performs better.

From results and results comparison with other literature on MWF19 dataset, the following conclusions could be drawn: First and foremost, no matter which options these eleven kinds of MWF are classified, approach 1 can distinguish them well. Then, when these eleven MWF is classified into 11 classes, approach 2 performs better than approach 1. Thirdly, specify to this dataset, reverse part data are most useful for data clustering. Furthermore, Under assigned settings, approach 3.2 distinguish reference fluid and test fluid 2-10 perfectly. Lastly, reasons for the non-ideal results from reference and fluid 1 are assumed from physical point of view.

From results of approach 1, 2.0, 3.2 and results comparison with other literature on MWF16 dataset, such conclusions could be drawn: Firstly, no matter how five kinds of MWF are classified, approach 1 can distinguish them well. Then, when these five kinds of MWF is classified into 5 classes, approach 2.0 performs not good as approach 1. This denotes that parameters and hyperparameters optimization has great influence on results. In addition, continuous wavelet transform is more suitable for feature extraction than STFT and HHT when specify to MWF16 dataset. Finally, data from reverse part are easier to cluster than data from forward part.

Based on results of these four datasets from proposed approaches, the following assumption can be drawn: First of all, data selection has significant impact on results. Secondly, for varied signals, suitable data processing are also different. Then, parameters in data processing and hyperparameters in ML algorithm has great significance on data distinction. In addition, besides time-frequency information analysis in scalogram, CWT can also be applied for boundaries searching in time domain. Finally, comparing with supervised learning, unsupervised learning approaches is harder to get good results.

7.3 Outlook

For the IS dataset, as the results from different data split way are different significantly, other data processing method and ML should be tried to find the best suitable option for EEG data classification. Besides, although the approaches 1 and 2 get good results on the selected datasets, they are not verified by datasets in other fields. For the future work, other datasets can be applied for validating the proposed approaches performance. Besides, at present approach 3.2 is just a framework, other data processing methods and ML algorithms can ge integrated into it. As time limitation, parameters and hyperparameters are not optimized in approach 3.2. For future work, parameters and hyperparameters can be optimized.

Bibliography

[AAL+21] An, Jing ; Ai, Ping ; Liu, Cong ; Xu, Sen ; Liu, Dakun: Deep clustering bearing fault diagnosis method based on local manifold learning of an autoencoded embedding. In: *IEEE Access* 9 (2021), pp. 30154–30168

[AAM15] Abdulkader, Sarah N. ; Atia, Ayman ; Mostafa, Mostafa-Sami M.: Brain computer interfacing: Applications and challenges. In: *Egyptian Informatics Journal* 16 (2015), Nr. 2, pp. 213–230

[AB18] Atkin, Christopher A. ; Badham, Stephen P.: The Phenomenological Influence of Inner Speech on Executive Functions. In: *Journal of Undergraduate Research at NTU* 1 (2018), Nr. 1, pp. 223–251

[ABS+19] Abiri, Reza ; Borhani, Soheil ; Sellers, Eric W. ; Jiang, Yang ; Zhao, Xiaopeng: A comprehensive review of EEG-based brain–computer interface paradigms. In: *Journal of neural engineering* 16 (2019), Nr. 1, pp. 011001

[ADC21] Ayana, Gelan ; Dese, Kokeb ; Choe, Se-woon: Transfer learning in breast cancer diagnoses via ultrasound imaging. In: *Cancers* 13 (2021), Nr. 4, pp. 738

[ADF15] Alderson-Day, Ben ; Fernyhough, Charles: Inner speech: development, cognitive functions, phenomenology, and neurobiology. In: *Psychological bulletin* 141 (2015), Nr. 5, pp. 931

[ADMW+18] Alderson-Day, Ben ; Mitrenga, Kaja ; Wilkinson, Sam ; McCarthy-Jones, Simon ; Fernyhough, Charles: The varieties of inner speech questionnaire–revised (VISQ-R): replicating and refining links between inner speech and psychopathology. In: *Consciousness and cognition* 65 (2018), pp. 48–58

[AFAS+20] Alzubaidi, Laith ; Fadhel, Mohammed A. ; Al-Shamma, Omran ; Zhang, Jinglan ; Santamaría, J ; Duan, Ye ; R. Oleiwi, Sameer: Towards a better understanding of transfer learning for medical imaging: a case study. In: *Applied Sciences* 10 (2020), Nr. 13, pp. 4523

[AGM16] Abhang, Priyanka A. ; Gawali, Bharti W. ; Mehrotra, Suresh C.: Technological basics of EEG recording and operation of apparatus. In: *Introduction to EEG-and Speech-Based Emotion Recognition* (2016), pp. 19–50

[AI22] AI, ARTIFICIAL I.: The fourth industrial revolution. In: *Presented at the American Psychological Association in Minneapolis* 4 (2022), pp. 6

[AN21] Ahmed, Latief ; Nabi, Firasath: AI (Artificial Intelligence) Driven Smart Agriculture. In: *Agriculture 5.0: Artificial Intelligence, IoT, and Machine Learning*. CRC Press, 2021, pp. 123–134

[ANHAOAW17] Al-Nafjan, Abeer ; Hosny, Manar ; Al-Ohali, Yousef ; Al-Wabil, Areej: Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review. In: *Applied Sciences* 7 (2017), Nr. 12, pp. 1239

[AOD+21] Angrick, Miguel ; Ottenhoff, Maarten C. ; Diener, Lorenz ; Ivucic, Darius ; Ivucic, Gabriel ; Goulis, Sophocles ; Saal, Jeremy ; Colon, Albert J. ; Wagner, Louis ; Krusienski, Dean J. [u. a.]: Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. In: *Communications biology* 4 (2021), Nr. 1, pp. 1–10

[AP06] Abdalla, HS ; Patel, S: The performance and oxidation stability of sustainable metalworking fluid derived from vegetable extracts. In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 220 (2006), Nr. 12, pp. 2027–2040

[APG15] Acharya, Shailesh ; Pant, Ashok K. ; Gyawali, Prashnna K.: Deep learning based large scale handwritten Devanagari character recognition. In: *2015 9th International conference on software, knowledge, information management and applications (SKIMA)* IEEE, 2015, pp. 1–6

[ARAS16] Acharya, Deepshikha ; Rani, Asha ; Agarwal, Shivangi ; Singh, Vijander: Application of adaptive Savitzky–Golay filter for EEG signal processing. In: *Perspectives in science* 8 (2016), pp. 677–679

[ARMH20] Atmani, Youcef ; Rechak, Said ; Mesloub, Ammar ; Hemmouche, Larbi: Enhancement in bearing fault classification parameters using Gaussian mixture models and mel frequency cepstral coefficients features. In: *Archives of Acoustics* 45 (2020), Nr. 2, pp. 283–295

[Ayo10] Ayodele, Taiwo O.: Types of machine learning algorithms. In: *New advances in machine learning* 3 (2010), pp. 19–48

[AZH+21] Alzubaidi, Laith ; Zhang, Jinglan ; Humaidi, Amjad J. ; Al-Dujaili, Ayad ; Duan, Ye ; Al-Shamma, Omran ; Santamaría, José ; Fadhel, Mohammed A. ; Al-Amidie, Muthana ; Farhan, Laith: Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. In: *Journal of big Data* 8 (2021), Nr. 1, pp. 1–74

[BAB14] Burns, Alexis ; Adeli, Hojjat ; Buford, John A.: Brain–computer interface after nervous system injury. In: *The Neuroscientist* 20 (2014), Nr. 6, pp. 639–651

[Bar11] Barnhart, Bradley L.: *The Hilbert-Huang transform: theory, applications, development*, The University of Iowa, Diss., 2011

[BAR20] Bathla, Gourav ; Aggarwal, Himanshu ; Rani, Rinkle: AutoTrustRec: Recommender system with social trust and deep learning using autoEncoder. In: *Multimedia Tools and Applications* 79 (2020), Nr. 29, pp. 20845–20860

[BB01] Baldi, Pierre ; Brunak, Søren: *Bioinformatics: the machine learning approach*. MIT press, 2001

[BC09] Brocklehurst, Paul ; Corley, Martin: Lexical bias and the phonemic similarity effect in inner speech. In: *15th Annual Conference on Architectures and Mechanisms for Language Processing*, 2009, pp. 7–9

[BDA21] Berg, Bram van d. ; Donkelaar, Sander van ; Alimardani, Maryam: Inner Speech Classification using EEG Signals: A Deep Learning Approach. In: *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)* IEEE, 2021, pp. 1–4

[BF00] Bruha, Ivan ; Famili, A: Postprocessing in machine learning and data mining. In: *ACM SIGKDD Explorations Newsletter* 2 (2000), Nr. 2, pp. 110–114

[BFM+08] Bierla, Aleksandra ; Fromentin, Guillaume ; Martin, J-M ; Le Mogne, Thierry ; Genet, Nicole: Tribological aspect of lubrication in form tapping of high-strength steel. In: *Lubrication Science* 20 (2008), Nr. 4, pp. 269–281

[BFR+98] Bradley, Paul S. ; Fayyad, Usama ; Reina, Cory [u. a.]: Scaling EM (expectation-maximization) clustering to large databases. In: *Microsoft Research* (1998), pp. 0–25

[BKS+11] Bermingham, MJ ; Kirsch, J ; Sun, Shoujin ; Palanisamy, S ; Dargusch, MS: New observations on tool life, cutting forces and chip morphology in cryogenic machining Ti-6Al-4V. In: *International Journal of Machine Tools and Manufacture* 51 (2011), Nr. 6, pp. 500–511

[BM98] Bloedorn, Eric ; Michalsi, RS: Data-driven constructive induction. In: *IEEE Intelligent Systems and their Applications* 13 (1998), Nr. 2, pp. 30–37

[BMDA16] Boudiaf, Adel ; Moussaoui, Abdelkrim ; Dahane, Amine ; Atoui, Issam: A comparative study of various methods of bearing faults diagnosis using the case Western Reserve University data. In: *Journal of Failure Analysis and Prevention* 16 (2016), Nr. 2, pp. 271–284

[BMHCH15] Brinksmeier, E ; Meyer, D ; Huesmann-Cordes, AG ; Herrmann, C: Metalworking fluids—mechanisms and performance. In: *CIRP Annals* 64 (2015), Nr. 2, pp. 605–628

[BN06] Bishop, Christopher M. ; Nasrabadi, Nasser M.: *Pattern recognition and machine learning*. Springer, 2006

[BNK20] Brunton, Steven L. ; Noack, Bernd R. ; Koumoutsakos, Petros: Machine learning for fluid mechanics. In: *Annual review of fluid mechanics* 52 (2020), pp. 477–508

[BRK+18] Bahaadini, Sara ; Rohani, Neda ; Katsaggelos, Aggelos K. ; Noroozi, Vahid ; Coughlin, Scott ; Zevin, Michael: Direct: Deep discriminative embedding for clustering of ligo data. In: *2018 25th ieee international conference on image processing (icip)* IEEE, 2018, pp. 748–752

[BS16] Bayar, Belhassen ; Stamm, Matthew C.: A deep learning approach to universal image manipulation detection using a new convolutional layer. In: *Proceedings of the 4th ACM workshop on information hiding and multimedia security*, 2016, pp. 5–10

[BYF+08] Bellini, Alberto ; Yazidi, Amine ; Filippetti, Fiorenzo ; Rossi, Claudio ; Capolino, Gerard-Andre: High frequency resolution techniques for rotor fault detection of induction machines. In: *IEEE Transactions on Industrial Electronics* 55 (2008), Nr. 12, pp. 4200–4209

[Can11] Canter, Neil: Monitoring metalworking fluids. In: *Tribology & lubrication technology* 67 (2011), Nr. 3, pp. 42

[CFC19] Cooney, Ciaran ; Folli, Raffaella ; Coyle, Damien: Optimizing layers improves CNN generalization and transfer learning for imagined speech decoding from EEG. In: *2019 IEEE international conference on systems, man and cybernetics (SMC)* IEEE, 2019, pp. 1311–1316

[CGGX02] Cheng, Ming ; Gao, Xiaorong ; Gao, Shangkai ; Xu, Dingfeng: Design and implementation of a brain-computer interface with high transfer rates. In: *IEEE transactions on biomedical engineering* 49 (2002), Nr. 10, pp. 1181–1186

[CGR17] Coretto, Germán A P. ; Gareis, Iván E ; Rufiner, H L.: Open access database of EEG signals recorded during imagined speech. In: *12th International Symposium on Medical Information Processing and Analysis* Bd. 10160 SPIE, 2017, pp. 1016002

[Chu18] Chu, Xiaojing: *The response mechanism of ecosystem CO2 exchange on precipitation distribution over a supra-tidal wetland in the Yellow River Delta*, Yantai Institute of Coastal Zone Research, University of Chinese Academy of Sciences, Diss., 2018

[CK10] Chandrasekaran, Bharath ; Kraus, Nina: The scalp-recorded brainstem response to speech: Neural origins and plasticity. In: *Psychophysiology* 47 (2010), Nr. 2, pp. 236–246

[CKFC20] Cooney, Ciaran ; Korik, Attila ; Folli, Raffaella ; Coyle, Damien: Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech EEG. In: *Sensors* 20 (2020), Nr. 16, pp. 4629

[CLZG20] Cooper, Clayton ; Liu, Dongdong ; Zhang, Jianjing ; Gao, Robert X.: Feature-Based Transfer Learning for Bearing Fault Recognition Without Available Fault Data. In: *International Symposium on Flexible Automation* Bd. 83617 American Society of Mechanical Engineers, 2020, pp. V001T08A004

[CS18] Chauhan, Nitin K. ; Singh, Krishna: A review on conventional machine learning vs deep learning. In: *2018 International conference on computing, power and communication technologies (GUCON)* IEEE, 2018, pp. 347–352

[DCR+92] Demonet, Jean-François ; Chollet, Francois ; Ramsay, Stuart ; Cardebat, Dominique ; Nespoulous, Jean-Luc ; Wise, Richard ; Rascol, André ; Frackowiak, Richard: The anatomy of phonological and semantic processing in normal subjects. In: *Brain* 115 (1992), Nr. 6, pp. 1753–1768

[DDPR15] Das, Sumit ; Dey, Aritra ; Pal, Akash ; Roy, Nabamita: Applications of artificial intelligence in machine learning: review and prospect. In: *International Journal of Computer Applications* 115 (2015), Nr. 9, pp. 31–41

[DFBW21] Dash, Debadatta ; Ferrari, Paul ; Berstis, Karinne ; Wang, Jun: Imagined, Intended, and Spoken Speech Envelope Synthesis from Neuromagnetic Signals. In: *International Conference on Speech and Computer* Springer, 2021, pp. 134–145

[DL05] Delmas, Robert ; Liu, Yan: Exploring student's conceptions of the standard deviation. In: *Statistics Education Research Journal* 4 (2005), Nr. 1, pp. 55–82

[Do49] Do, Hebb: The organization of behavior. In: *New York* 50 (1949), pp. 437

[Dro00] Dronkers, Nina F.: The pursuit of brain–language relationships. In: *Brain and Language* 71 (2000), Nr. 1, pp. 59–61

[DS20] Demmerling, AL ; Söffker, D: Improved examination and test procedure of tapping torque tests according to ASTM D5619 using coated forming taps and water-mixed metalworking fluids. In: *Tribology International* 145 (2020), pp. 106151

[DSD+21] Dinic, Filip ; Singh, Kamalpreet ; Dong, Tony ; Rezazadeh, Milad ; Wang, Zhibo ; Khosrozadeh, Ali ; Yuan, Tiange ; Voznyy, Oleksandr: Applied Machine Learning for Developing Next-Generation

Functional Materials. In: *Advanced Functional Materials* 31 (2021), Nr. 51, pp. 2104195

[DSW+18] Dong, Xingping ; Shen, Jianbing ; Wang, Wenguan ; Liu, Yu ; Shao, Ling ; Porikli, Fatih: Hyperparameter optimization for tracking with continuous deep q-learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 518–527

[DTS+19] Dai, Juying ; Tang, Jian ; Shao, Faming ; Huang, Shuzhan ; Wang, Yangyang: Fault diagnosis of rolling bearing based on multiscale intrinsic mode function permutation entropy and a stacked sparse denoising autoencoder. In: *Applied Sciences* 9 (2019), Nr. 13, pp. 2743

[DW+18] Di, Jian ; Wang, Leilei [u. a.]: Application of improved deep auto-encoder network in rolling bearing fault diagnosis. In: *Journal of Computer and Communications* 6 (2018), Nr. 07, pp. 41

[DWVVJ+04] Dronkers, Nina F. ; Wilkins, David P. ; Van Valin Jr, Robert D. ; Redfern, Brenda B. ; Jaeger, Jeri J.: Lesion analysis of the brain areas involved in language comprehension. In: *Cognition* 92 (2004), Nr. 1-2, pp. 145–177

[ECS11] Erisoglu, Murat ; Calis, Nazif ; Sakallioglu, Sadullah: A new algorithm for initial cluster centers in k-means algorithm. In: *Pattern Recognition Letters* 32 (2011), Nr. 14, pp. 1701–1705

[ELM20] Elmachtoub, Adam ; Liang, Jason Cheuk N. ; McNellis, Ryan: Decision trees for decision-making under the predict-then-optimize framework. In: *International Conference on Machine Learning* PMLR, 2020, pp. 2858–2867

[EM03] Emerson, Michael J. ; Miyake, Akira: The role of inner speech in task switching: A dual-task investigation. In: *Journal of Memory and Language* 48 (2003), Nr. 1, pp. 148–168

[ESS05] Eisenberg, Nancy ; Sadovsky, Adrienne ; Spinrad, Tracy L.: Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. In: *New directions for child and adolescent development* 2005 (2005), Nr. 109, pp. 109–118

[FHMO04] Fujimaki, Norio ; Hayakawa, Tomoe ; Matani, Ayumu ; Okabe, Yoichi: Right-lateralized neural activity during inner speech repeated by cues. In: *Neuroreport* 15 (2004), Nr. 15, pp. 2341–2345

[FLXL16] Fu, Sheng ; Liu, Kun ; Xu, Yonggang ; Liu, Yi: Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy-means clustering. In: *Shock and Vibration* 2016 (2016), pp. 1–8

[Fos20] Fossa, Pablo: The representational and the expressive: Two functions of the inner speech. In: *Psicologia: Teoria e Pesquisa* 35 (2020)

[Fra20] Fradkov, Alexander L.: Early history of machine learning. In: *IFAC-PapersOnLine* 53 (2020), Nr. 2, pp. 1385–1390

[GBK+13] Groppe, David M. ; Bickel, Stephan ; Keller, Corey J. ; Jain, Sanjay K. ; Hwang, Sean T. ; Harden, Cynthia ; Mehta, Ashesh D.: Dominant frequencies of resting human brain activity as measured by the electrocorticogram. In: *Neuroimage* 79 (2013), pp. 223–233

[GC11] Goswami, Jaideva C. ; Chan, Andrew K.: *Fundamentals of wavelets: theory, algorithms, and applications*. John Wiley & Sons, 2011

[GCTGRGVP17] González-Castañeda, Erick F. ; Torres-García, Alejandro A. ; Reyes-García, Carlos A. ; Villaseñor-Pineda, Luis: Sonification and textification: Proposing methods for classifying unspoken words from EEG signals. In: *Biomedical Signal Processing and Control* 37 (2017), pp. 82–91

[GD05] Grauman, Kristen ; Darrell, Trevor: The pyramid match kernel: Discriminative classification with sets of image features. In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* Bd. 2 IEEE, 2005, pp. 1458–1465

[GDS19] Ghosh, Sourish ; Dasgupta, Anasuya ; Swetapadma, Aleena: A study on support vector machine based linear and non-linear pattern classification. In: *2019 International Conference on Intelligent Sustainable Systems (ICISS)* IEEE, 2019, pp. 24–28

[GKMJ22] Greener, Joe G. ; Kandathil, Shaun M. ; Moffat, Lewis ; Jones, David T.: A guide to machine learning for biologists. In: *Nature Reviews Molecular Cell Biology* 23 (2022), Nr. 1, pp. 40–55

[GLB15] Gholizadeh, Samira ; Leman, Z ; Baharudin, BT Hang T.: A review of the application of acoustic emission technique in engineering. In: *Structural Engineering and Mechanics* 54 (2015), Nr. 6, pp. 1075–1095

[GP16] Goyal, D ; Pabla, BS: The vibration monitoring methods and signal processing techniques for structural health monitoring: a review. In: *Archives of Computational Methods in Engineering* 23 (2016), Nr. 4, pp. 585–594

[Gre80] Green, Robert E.: Basic wave analysis of acoustic emission. In: *Mechanics of nondestructive testing*. Springer, 1980, pp. 55–76

[GSA+18] Golas, Sara B. ; Shibahara, Takuma ; Agboola, Stephen ; Otaki, Hiroko ; Sato, Jumpei ; Nakae, Tatsuya ; Hisamitsu, Toru ; Kojima, Go ; Felsted, Jennifer ; Kakarmath, Sujay: A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. In: *BMC medical informatics and decision making* 18 (2018), Nr. 1, pp. 1–17

[GT14] Garg, Dweepna ; Trivedi, Khushboo: Fuzzy k-mean clustering in mapreduce on cloud based hadoop. In: *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* IEEE, 2014, pp. 1607–1610

[HAMI17] Hashim, Noramiza ; Ali, Aziah ; Mohd-Isa, Wan-Noorshahida: Word-based classification of imagined speech using EEG. In: *International Conference on Computational Science and Technology* Springer, 2017, pp. 195–204

[HBS+17] Huppert, Theodore ; Barker, Jeff ; Schmidt, Benjamin ; Walls, Shawn ; Ghuman, Avniel: Comparison of group-level, source localized activity for simultaneous functional near-infrared spectroscopy-magnetoencephalography and simultaneous fNIRS-fMRI during parametric median nerve stimulation. In: *Neurophotonics* 4 (2017), Nr. 1, pp. 015001

[HCMBS14] Huesmann-Cordes, AG ; Meyer, D ; Brinksmeier, E ; Schulz, J: Influence of additives in metalworking fluids on the wear resistance of steels. In: *Procedia CIRP* 13 (2014), pp. 108–113

[HHSH91] Hitch, Graham J. ; Halliday, M S. ; Schaafstal, Alma M. ; Heffernan, Thomas M.: Speech,"inner speech," and the development of short-term memory: Effects of picture-labeling on recall. In: *Journal of Experimental Child Psychology* 51 (1991), Nr. 2, pp. 220–234

[HKCI13] Hwang, Han-Jeong ; Kim, Soyoun ; Choi, Soobeom ; Im, Chang-Hwan: EEG-based brain-computer interfaces: a thorough literature survey. In: *International Journal of Human-Computer Interaction* 29 (2013), Nr. 12, pp. 814–826

[HOT06] Hinton, Geoffrey E. ; Osindero, Simon ; Teh, Yee-Whye: A fast learning algorithm for deep belief nets. In: *Neural computation* 18 (2006), Nr. 7, pp. 1527–1554

[Hua19] Hua, Feng: Rolling Bearing Anomaly Detection Based on Generative Adversarial Networks. In: *Artificial intelligence and robotics research* 8 (2019), Nr. 4, pp. 208–218

[HZZ+19] Huang, Lei ; Zhou, Yi ; Zhu, Fan ; Liu, Li ; Shao, Ling: Iterative normalization: Beyond standardization towards efficient whitening. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4874–4883

[ISN+18] Injadat, MohammadNoor ; Salo, Fadi ; Nassif, Ali B. ; Essex, Aleksander ; Shami, Abdallah: Bayesian optimization with machine learning algorithms towards anomaly detection. In: *2018 IEEE global communications conference (GLOBECOM)* IEEE, 2018, pp. 1–6

[JF07] Jones, Simon R. ; Fernyhough, Charles: Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. In: *Clinical psychology review* 27 (2007), Nr. 2, pp. 140–154

[JG14] Jani, Mansi P. ; Gore, Geeta B.: Occurrence of communication and swallowing problems in neurological disorders: analysis of forty patients. In: *NeuroRehabilitation* 35 (2014), Nr. 4, pp. 719–727

[JHZ+19] Jiang, Wenqian ; Hong, Yang ; Zhou, Beitong ; He, Xin ; Cheng, Cheng: A GAN-based anomaly detection approach for imbalanced industrial time series. In: *IEEE Access* 7 (2019), pp. 143608–143619

[JK20] Jakhar, Deepack ; Kaur, Ishmeet: Artificial intelligence, machine learning and deep learning: definitions and differences. In: *Clinical and experimental dermatology* 45 (2020), Nr. 1, pp. 131–132

[JNR05] Jain, Anil ; Nandakumar, Karthik ; Ross, Arun: Score normalization in multimodal biometric systems. In: *Pattern recognition* 38 (2005), Nr. 12, pp. 2270–2285

[Joh99] Johansson, Mathias: *The hilbert transform*, Växjö University, Diss., 1999

[KAF+08] Keim, Daniel ; Andrienko, Gennady ; Fekete, Jean-Daniel ; Görg, Carsten ; Kohlhammer, Jörn ; Melançon, Guy: Visual analytics: Definition, process, and challenges. In: *Information visualization*. Springer, 2008, pp. 154–175

[Kay04] Kaya, Ahmet: Outlier effects on databases. In: *International Conference on Advances in Information Systems* Springer, 2004, pp. 88–95

[KCJ09] Khalily, Muhammad T. ; Clarke, Liam ; Jahangir, Farhana: Alpha-theta Neurofeedback therapy: A non-invasive treatment for addiction and post-Traumatic Stress disorder. In: *FWU Journal of Social Sciences* 3 (2009), Nr. 2, pp. 83

[KGL+17] Kljajevic, Vanja ; Gómez, Estibaliz U. ; López, Cristina ; Bandeira, Yolanda B. ; Vicente, Agustin: Inner speech in post-stroke motor aphasia. In: *CogSci*, 2017

[KMC+21] Kamat, Pooja ; Marni, Pallavi ; Cardoz, Lester ; Irani, Arshan ; Gajula, Anuj ; Saha, Akash ; Kumar, Satish ; Sugandhi, Rekha: Bearing Fault Detection Using Comparative Analysis of Random Forest, ANN, and Autoencoder Methods. In: *Communication and Intelligent Systems*. Springer, 2021, pp. 157–171

[KSG+18] Kamavuako, Ernest N. ; Sheikh, Usman A. ; Gilani, Syed O. ; Jamil, Mohsin ; Niazi, Imran K.: Classification of overt and covert speech for near-infrared spectroscopy-based brain computer interface. In: *Sensors* 18 (2018), Nr. 9, pp. 2989

[KT18] Kang, Myeongsu ; Tian, Jing: Machine Learning: Data Pre-processing. In: *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things* (2018), pp. 111–130

[Lan19] Lantz, Brett: *Machine learning with R: expert techniques for predictive modeling*. Packt publishing ltd, 2019

[LBB+14] Lauro, CH ; Brandão, LC ; Baldo, D ; Reis, RA ; Davim, JP: Monitoring and processing signal applied in machining processes–A review. In: *Measurement* 58 (2014), pp. 73–86

[LBC+18] Lotte, Fabien ; Bougrain, Laurent ; Cichocki, Andrzej ; Clerc, Maureen ; Congedo, Marco ; Rakotomamonjy, Alain ; Yger, Florian: A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. In: *Journal of neural engineering* 15 (2018), Nr. 3, pp. 031005

[LBM+18] Liakos, Konstantinos G. ; Busato, Patrizia ; Moshou, Dimitrios ; Pearson, Simon ; Bochtis, Dionysis: Machine learning in agriculture: A review. In: *Sensors* 18 (2018), Nr. 8, pp. 2674

[LCL+07] Lotte, Fabien ; Congedo, Marco ; Lécuyer, Anatole ; Lamarche, Fabrice ; Arnaldi, Bruno: A review of classification algorithms for EEG-based brain–computer interfaces. In: *Journal of neural engineering* 4 (2007), Nr. 2, pp. R1

[LCN12] Lawal, Sunday A. ; Choudhury, Imtiaz A. ; Nukman, Yusoff: Application of vegetable oil-based metalworking fluids in machining ferrous metals—a review. In: *International Journal of Machine Tools and Manufacture* 52 (2012), Nr. 1, pp. 1–12

[LFMW09] Lidstone, Jane S. ; Fernyhough, Charles ; Meins, Elizabeth ; Whitehouse, Andrew J.: Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. In: *Journal of autism and developmental disorders* 39 (2009), Nr. 8, pp. 1222–1225

[LHCT18] Lin, Tang-Huang ; Hsiao, Min-Chung ; Chan, Hai-Po ; Tsai, Fuan: A Novel Approach to Relative Radiometric Calibration on Spatial and Temporal Variations for FORMOSAT-5 RSI Imagery. In: *Sensors* 18 (2018), Nr. 7, pp. 1996

[LHV18] Langland-Hassan, Peter ; Vicente, Agustín: *Inner speech: New voices*. Oxford University Press, USA, 2018

[Lig21] Light, Glenn: Nondestructive evaluation technologies for monitoring corrosion. In: *Techniques for Corrosion Monitoring*. Elsevier, 2021, pp. 285–304

[LIS16] Leonetti, Matteo ; Iocchi, Luca ; Stone, Peter: A synthesis of automated planning and reinforcement learning for efficient, robust decision-making. In: *Artificial Intelligence* 241 (2016), pp. 103–130

[LLL18] Liu, Guozeng ; Li, Haiping ; Liu, Wei: Bearing fault detection in varying operational conditions based on empirical mode decomposition and random forest. In: *2018 Prognostics and System Health Management Conference (PHM-Chongqing)* IEEE, 2018, pp. 851–854

[LLT+17] Li, Shaobo ; Liu, Guokai ; Tang, Xianghong ; Lu, Jianguang ; Hu, Jianjun: An ensemble deep convolutional neural network model with improved DS evidence fusion for bearing fault diagnosis. In: *Sensors* 17 (2017), Nr. 8, pp. 1729

[LMA+16] Laurent, Lucie ; Millot, Jean-Louis ; Andrieu, Patrice ; Camos, Valérie ; Floccia, Caroline ; Mathy, Fabien: Inner speech sustains predictable task switching: direct evidence in adults. In: *Journal of Cognitive Psychology* 28 (2016), Nr. 5, pp. 585–592

[LMF10] Lidstone, Jane S. ; Meins, Elizabeth ; Fernyhough, Charles: The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. In: *Journal of Experimental Child Psychology* 107 (2010), Nr. 4, pp. 438–451

[LSSG17] Liang, Hong ; Sun, Xiao ; Sun, Yunlei ; Gao, Yuan: Text feature extraction based on deep learning: a review. In: *EURASIP journal on wireless communications and networking* 2017 (2017), Nr. 1, pp. 1–12

[LT06] Lukin, Alexey ; Todd, Jeremy: Adaptive time-frequency resolution for analysis and processing of audio. In: *Audio Engineering Society Convention 120* Audio Engineering Society, 2006, pp. 1–10

[LVD21] Larocque-Villiers, Justin ; Dumond, Patrick: Towards Generalization of Intelligent Fault Detection for Roller Element Bearings via Distinct Dataset Transfer Learning. In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* Bd. 85475 American Society of Mechanical Engineers, 2021, pp. V010T10A012

[LWDZ18] Liang, Tianchen ; Wu, Shuaipeng ; Duan, Wenjing ; Zhang, Ruishan: Bearing fault diagnosis based on improved ensemble learning and deep belief network. In: *Journal of Physics: Conference Series* Bd. 1074 IOP Publishing, 2018, pp. 012154

[LWSH19] Li, Yongbo ; Wang, Xianzhi ; Si, Shubin ; Huang, Shiqian: Entropy based fault classification using the Case Western Reserve University data: A benchmark study. In: *IEEE Transactions on Reliability* 69 (2019), Nr. 2, pp. 754–767

[LZLL15] Liao, Guangchun ; Zhu, Haiping ; Liu, Kangjun ; Liao, JiaWei: Fault diagnosis of rolling bearing based on PSO and continuous Gaussian mixture HMM. In: *2015 2nd International Conference on Machinery, Materials Engineering, Chemical Engineering and Biotechnology* Atlantis Press, 2015, pp. 911–917

[LZM12] Li, Wenjie ; Zhao, Jiankang ; Meng, Shaoxin: Partial discharge time-frequency spectrum analysis and extraction for power cable. In: *2012 Asia-Pacific Power and Energy Engineering Conference* IEEE, 2012, pp. 1–3

[LZMZ18] Li, Yupeng ; Zhang, Jianhua ; Ma, Zhanyu ; Zhang, Yu: Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model. In: *IEEE Transactions on Big Data* 6 (2018), Nr. 2, pp. 223–232

[LZT95] Li, Cuiwei ; Zheng, Chongxun ; Tai, Changfeng: Detection of ECG characteristic points using wavelet transforms. In: *IEEE Transactions on biomedical Engineering* 42 (1995), Nr. 1, pp. 21–28

[Mah20] Mahesh, Batta: Machine learning algorithms-a review. In: *International Journal of Science and Research (IJSR)* 9 (2020), pp. 381–386

[Man96] Manhart, Klaus: Artificial intelligence modelling: Data driven and theory driven approaches. In: *Social Science Microsimulation*. Springer, 1996, pp. 416–431

[Mar91] Markov, Zdravko: An approach to data-driven learning. In: *International Workshop on Fundamentals of Artificial Intelligence Research* Springer, 1991, pp. 127–140

[MBD+90] Mitchell, Tom ; Buchanan, Bruce ; DeJong, Gerald ; Dietterich, Thomas ; Rosenbloom, Paul ; Waibel, Alex: Machine learning. In: *Annual review of computer science* 4 (1990), Nr. 1, pp. 417–433

[McC94] McCoy, JS: Tracing the Historical Development of Metalworking Fluids. In: *Manufacturing engineering and materials processing* 41 (1994), pp. 1–1

[MDS07] Mesgarani, Nima ; David, Stephen ; Shamma, Shihab: Representation of phonemes in primary auditory cortex: how the brain analyzes speech. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* Bd. 4 IEEE, 2007, pp. IV–765

[MGB+11] Myint, Soe W. ; Gober, Patricia ; Brazel, Anthony ; Grossman-Clarke, Susanne ; Weng, Qihao: Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. In: *Remote sensing of environment* 115 (2011), Nr. 5, pp. 1145–1161

[MIM+18] Martin, Stephanie ; Iturrate, Iñ. ; Millán, José del R ; Knight, Robert T. ; Pasley, Brian N.: Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis. In: *Frontiers in neuroscience* 12 (2018), pp. 422

[Mit06] Mitchell, Tom M.: *The discipline of machine learning*. Carnegie Mellon University, School of Computer Science, Machine Learning, 2006

[MJ08] Marksberry, PW ; Jawahir, IS: A comprehensive tool-wear/tool-life performance model in the evaluation of NDM (near dry machining) for sustainable manufacturing. In: *International Journal of Machine Tools and Manufacture* 48 (2008), Nr. 7-8, pp. 878–886

[MJJL08] Mallapragada, Pavan K. ; Jin, Rong ; Jain, Anil K. ; Liu, Yi: Semiboost: Boosting for semi-supervised learning. In: *IEEE transactions on pattern analysis and machine intelligence* 31 (2008), Nr. 11, pp. 2000–2014

[MMM+16] Mahoney, Vanessa M. ; Mezzano, Valeria ; Mirams, Gary R. ; Maass, Karen ; Li, Zhen ; Cerrone, Marina ; Vasquez, Carolina ; Bapat, Aneesh ; Delmar, Mario ; Morley, Gregory E.: Connexin43 contributes to electrotonic conduction across scar tissue in the intact heart. In: *Scientific reports* 6 (2016), Nr. 1, pp. 1–12

[MOM+17] Moffett, Steven X. ; O'Malley, Sean M. ; Man, Shushuang ; Hong, Dawei ; Martin, Joseph V.: Dynamics of high frequency brain activity. In: *Scientific reports* 7 (2017), Nr. 1, pp. 1–5

[MP43] McCulloch, Warren S. ; Pitts, Walter: A logical calculus of the ideas immanent in nervous activity. In: *The bulletin of mathematical biophysics* 5 (1943), Nr. 4, pp. 115–133

[MP69] Minsky, Marvin ; Papert, Seymour: An introduction to computational geometry. In: *Cambridge tiass, HIT* 479 (1969), pp. 480

[MWF18] Maxwell, Aaron E. ; Warner, Timothy A. ; Fang, Fang: Implementation of machine-learning classification in remote sensing: An applied review. In: *International Journal of Remote Sensing* 39 (2018), Nr. 9, pp. 2784–2817

[NBP09] Nguyen, Giang H. ; Bouzerdoum, Abdesselam ; Phung, Son L.: Learning pattern classification tasks with imbalanced data sets. In: *Pattern recognition* (2009), pp. 193–208

[NKA17] Nguyen, Chuong H. ; Karavas, George K. ; Artemiadis, Panagiotis: Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. In: *Journal of neural engineering* 15 (2017), Nr. 1, pp. 016002

[Noo05] Nooteboom, Sieb G.: Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. In: *Speech communication* 47 (2005), Nr. 1-2, pp. 43–58

[NPR+22] Nieto, Nicolás ; Peterson, Victoria ; Rufiner, Hugo L. ; Kamienkowski, Juan E. ; Spies, Ruben: Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. In: *Scientific Data* 9 (2022), Nr. 1, pp. 1–17

[NRBZ15] Niknejad, Milad ; Rabbani, Hossein ; Babaie-Zadeh, Massoud: Image restoration using Gaussian mixture models with spatially constrained patch clustering. In: *IEEE Transactions on Image Processing* 24 (2015), Nr. 11, pp. 3624–3636

[OC03] Oaksford, Mike ; Chater, Nick: Optimal data selection: Revision, review, and reevaluation. In: *Psychonomic Bulletin & Review* 10 (2003), Nr. 2, pp. 289–318

[OCBR18] Ongsulee, Pariwat ; Chotchaung, Veena ; Bamrungsi, Eak ; Rodcheewit, Thanaporn: Big data, predictive analytics and machine learning. In: *2018 16th international conference on ICT and knowledge engineering (ICT&KE)* IEEE, 2018, pp. 1–6

[OJJ+22] Ong, Chiu L. ; Jiang, Xuefeng ; Juan, Joon C. ; Khaligh, Nader G. ; Heidelberg, Thorsten: Ashless and non-corrosive disulfide compounds as excellent extreme pressure additives in naphthenic oil. In: *Journal of Molecular Liquids* 351 (2022), pp. 118553

[Ong17] Ongsulee, Pariwat: Artificial intelligence, machine learning and deep learning. In: *2017 15th international conference on ICT and knowledge engineering (ICT&KE)* IEEE, 2017, pp. 1–6

[PBMT21] Pouyap, Mireille ; Bitjoka, Laurent ; Mfoumou, Etienne ; Toko, Denis: Improved Bearing Fault Diagnosis by Feature Extraction Based on GLCM, Fusion of Selection Methods, and Multiclass-Naïve Bayes Classification. In: *Journal of Signal and Information Processing* 12 (2021), Nr. 4, pp. 71–85

[PBRL+14] Perrone-Bertolotti, Marcela ; Rapin, Lucile ; Lachaux, J-P ; Baciu, Monica ; Loevenbruck, Hélène: What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. In: *Behavioural brain research* 261 (2014), pp. 220–239

[PCZ18] Pan, Tongvang ; Chen, Jinglong ; Zhou, Zitong: Intelligent fault diagnosis of rolling bearing via deep-layerwise feature extraction using deep belief network. In: *2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)* IEEE, 2018, pp. 509–514

[PLB+18] Požar, Tomaž ; Laloš, Jernej ; Babnik, Aleš ; Petkovšek, Rok ; Bethune-Waddell, Max ; Chau, Kenneth J. ; Lukasievicz, Gustavo V. ; Astrath, Nelson G.: Isolated detection of elastic waves driven by the momentum of light. In: *Nature communications* 9 (2018), Nr. 1, pp. 1–11

[PMM20] Padarian, José ; Minasny, Budiman ; McBratney, Alex B.: Machine learning and soil sciences: A review aided by machine learning tools. In: *Soil* 6 (2020), Nr. 1, pp. 35–52

[PP11] Papadatou-Pastou, Marietta: Handedness and language lateralization: Why are we right-handed and left-brained. In: *Hellenic Journal of Psychology* 8 (2011), Nr. 2, pp. 248–265

[PR21] Panachakel, Jerrin T. ; Ramakrishnan, Angarai G.: Decoding covert speech from EEG-a comprehensive review. In: *Frontiers in Neuroscience* (2021), pp. 392

[PSJL17] Pang, Yanwei ; Sun, Manli ; Jiang, Xiaoheng ; Li, Xuelong: Convolution in convolution for network in network. In: *IEEE transactions on neural networks and learning systems* 29 (2017), Nr. 5, pp. 1587–1597

[PWC+21] Pei, Yuhang ; Wei, Minxiang ; Chen, Kai ; Chen, Xinda ; Li, Shunming: A Fault Diagnosis Method of Rolling Bearing Based on CEEMDAN and KFCM. In: *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)* IEEE, 2021, pp. 1–6

[QGL+19] Qu, Kaiyang ; Guo, Fei ; Liu, Xiangrong ; Lin, Yuan ; Zou, Quan: Application of machine learning in microbiology. In: *Frontiers in microbiology* 10 (2019), pp. 827

[QWW+19] Qiao, Huihui ; Wang, Taiyong ; Wang, Peng ; Zhang, Lan ; Xu, Mingda: An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions. In: *Ieee Access* 7 (2019), pp. 118954–118964

[QZ11] Qiang, Wang ; Zhongli, Zhan: Reinforcement learning model, algorithms and its application. In: *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)* IEEE, 2011, pp. 1143–1146

[RABV14] Rauber, Thomas W. ; Assis Boldt, Francisco de ; Varejão, Flávio M.: Heterogeneous feature models and feature selection applied to bearing fault diagnosis. In: *IEEE Transactions on Industrial Electronics* 62 (2014), Nr. 1, pp. 637–646

[RDC20] Roy, Sayanjit S. ; Dey, Sayantan ; Chatterjee, Soumya: Autocorrelation aided random forest classifier-based bearing fault detection framework. In: *IEEE Sensors Journal* 20 (2020), Nr. 18, pp. 10792–10800

[Rey09] Reynolds, Douglas A.: Gaussian mixture models. In: *Encyclopedia of biometrics* 741 (2009)

[RM13] Rastogi, Nidhi ; Mehra, Rajesh: Analysis of Savitzky-Golay filter for baseline wander cancellation in ECG using wavelets. In: *Int. J. Eng. Sci. Emerg. Technol* 6 (2013), Nr. 1, pp. 2231–6604

[Ros58] Rosenblatt, Frank: The perceptron: a probabilistic model for information storage and organization in the brain. In: *Psychological review* 65 (1958), Nr. 6, pp. 386

[RS16] Rajasekar, Shanmuganathan ; Sanjuan, Miguel A.: *Nonlinear resonances*. Springer, 2016

[RSYW+19] Rezazadeh Sereshkeh, Alborz ; Yousefi, Rozhin ; Wong, Andrew T. ; Rudzicz, Frank ; Chau, Tom: Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech. In: *Brain-Computer Interfaces* 6 (2019), Nr. 4, pp. 128–140

[Rud19] Ruder, Sebastian: *Neural transfer learning for natural language processing*, NUI Galway, Diss., 2019

[SA18] Saini, Manish K. ; Aggarwal, Akanksha: Detection and diagnosis of induction motor bearing faults using multiwavelet transform and naive Bayes classifier. In: *International Transactions on Electrical Energy Systems* 28 (2018), Nr. 8, pp. e2577

[SAF15] Saidi, Lotfi ; Ali, Jaouher B. ; Fnaiech, Farhat: Application of higher order spectral features and support vector machines for bearing faults classification. In: *ISA transactions* 54 (2015), pp. 193–206

[Sch99] Schneider, David I.: *An introduction to programming using Visual Basic 6.0*. Prentice Hall Upper Saddle River, NJ, 1999

[Sch11] Schafer, Ronald W.: What is a Savitzky-Golay filter?[lecture notes]. In: *IEEE Signal processing magazine* 28 (2011), Nr. 4, pp. 111–117

[SG64] Savitzky, Abraham ; Golay, Marcel J.: Smoothing and differentiation of data by simplified least squares procedures. In: *Analytical chemistry* 36 (1964), Nr. 8, pp. 1627–1639

[SG02] Strehl, Alexander ; Ghosh, Joydeep: Cluster ensembles—a knowledge reuse framework for combining multiple partitions. In: *Journal of machine learning research* 3 (2002), Nr. Dec, pp. 583–617

[SI15] Surangsrirat, Decho ; Intarapanich, Apichart: Analysis of the meditation brainwave from consumer EEG device. In: *SoutheastCon 2015* IEEE, 2015, pp. 1–6

[SJLL18] Shao, Haidong ; Jiang, Hongkai ; Lin, Ying ; Li, Xingqiu: A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. In: *Mechanical Systems and Signal Processing* 102 (2018), pp. 278–297

[SJWW17] Shao, Haidong ; Jiang, Hongkai ; Wang, Fuan ; Wang, Yanan: Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. In: *ISA transactions* 69 (2017), pp. 187–201

[SK17] Sharma, Diksha ; Kumar, Neeraj: A review on machine learning algorithms, tasks and applications. In: *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 6 (2017), Nr. 10, pp. 2278–1323

[SK18] Schrider, Daniel R. ; Kern, Andrew D.: Supervised machine learning for population genetics: a new paradigm. In: *Trends in Genetics* 34 (2018), Nr. 4, pp. 301–312

[SK19] Sublime, Jérémie ; Kalinicheva, Ekaterina: Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami. In: *Remote Sensing* 11 (2019), Nr. 9, pp. 1123

[SKS16] Sun, Jun ; Kunegis, Jérôme ; Staab, Steffen: Predicting user roles in social networks using transfer learning with feature transformation. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* IEEE, 2016, pp. 128–135

[SLS17] Shah, Nikhil C. ; Lahiri, Srabona ; Sen, Dhrubo J.: Brainwaves as neurophysiological outcomes in neurosedation. In: *Pharma Science Monitor* 8 (2017), Nr. 2, pp. 306–337

[Smi08] Smith, Graham T.: *Cutting tool technology: industrial handbook*. Springer Science & Business Media, 2008

[SRD22] Schmid, Michael ; Rath, David ; Diebold, Ulrike: Why and How Savitzky–Golay Filters Should Be Replaced. In: *ACS measurement science Au* 2 (2022), Nr. 2, pp. 185–196

[SS14] Song, YoungJae ; Sepulveda, Francisco: Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state. In: *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings* IEEE, 2014, pp. 568–571

[SS18] Saravanan, R ; Sujatha, Pothula: A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In: *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* IEEE, 2018, pp. 945–949

[SS20] Singh, Dalwinder ; Singh, Birmohan: Investigating the impact of data normalization on classification performance. In: *Applied Soft Computing* 97 (2020), pp. 105524

[Sun13] Sun, Shiliang: A survey of multi-view machine learning. In: *Neural computing and applications* 23 (2013), Nr. 7, pp. 2031–2038

[SW83] Scruby, CB ; Wadley, HNG: An assessment of acoustic emission for nuclear pressure vessel monitoring. In: *Progress in Nuclear Energy* 11 (1983), Nr. 3, pp. 275–297

[T+95] Tesauro, Gerald [u. a.]: Temporal difference learning and TD-Gammon. In: *Communications of the ACM* 38 (1995), Nr. 3, pp. 58–68

[T+02] Teplan, Michal [u. a.]: Fundamentals of EEG measurement. In: *Measurement science review* 2 (2002), Nr. 2, pp. 1–11

[TA01] Terchi, A ; Au, YHJ: Acoustic emission signal processing. In: *Measurement and Control* 34 (2001), Nr. 8, pp. 240–244

[TFS+06] Terumitsu, Makoto ; Fujii, Yukihiko ; Suzuki, Kiyotaka ; Kwee, Ingrid L. ; Nakada, Tsutomu: Human primary motor cortex shows hemispheric specialization for speech. In: *Neuroreport* 17 (2006), Nr. 11, pp. 1091–1095

[Tra13] Trafny, Elżbieta A: Microorganisms in metalworking fluids: current issues in research and management. In: *International journal of occupational medicine and environmental health* 26 (2013), Nr. 1, pp. 4–15

[TS09] Taylor, Matthew E. ; Stone, Peter: Transfer learning for reinforcement learning domains: A survey. In: *Journal of Machine Learning Research* 10 (2009), Nr. 7, pp. 1–53

[TW04] Tan, Ying ; Wang, Jun: A support vector machine with a hybrid kernel and minimal Vapnik-Chervonenkis dimension. In: *IEEE Transactions on knowledge and data engineering* 16 (2004), Nr. 4, pp. 385–395

[VBR20] Varolgüneş, Yasemin B. ; Bereau, Tristan ; Rudzinski, Joseph F.: Interpretable embeddings from molecular simulations using Gaussian mixture variational autoencoders. In: *Machine Learning: Science and Technology* 1 (2020), Nr. 1, pp. 015012

[VGK17] Vakharia, V ; Gupta, VK ; Kankar, PK: Efficient fault diagnosis of ball bearing using ReliefF and Random Forest classifier. In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 39 (2017), Nr. 8, pp. 2969–2982

[VL63] Vapnik, VN ; Lerner, A Y.: Recognition of patterns with help of generalized portraits. In: *Avtomat. i Telemekh* 24 (1963), Nr. 6, pp. 774–780

[VPB13] Vercueil, Laurent ; Perronne-Bertolotti, Marcela: Ictal inner speech jargon. In: *Epilepsy & Behavior* 27 (2013), Nr. 2, pp. 307–309

[VS20] Viveka, P ; Selvi, F Kurus M.: A Modified K-Means Algorithm for Big Data Clustering. In: *Group* 1 (2020), Nr. 1, pp. 1–0

[VSP+22] Vishwendra, More A. ; Salunkhe, Pratiksha S. ; Patil, Shivanjali V. ; Shinde, Sumit A. ; Shinde, PV ; Desavale, RG ; Jadhav, PM ; Dharwadkar, Nagaraj V.: A Novel Method to Classify Rolling Element Bearing Faults Using K-Nearest Neighbor Machine Learning Algorithm. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 8 (2022), Nr. 3, pp. 031202

[WBH12] Winter, Marius ; Bock, Ralf ; Herrmann, Christoph: Investigation of a new ecologically benign metalworking fluid in abrasive machining processes to substitute mineral oil based fluids. In: *Procedia CIRP* 1 (2012), pp. 393–398

[WCJ17] Wu, Chenxi ; Chen, Tefang ; Jiang, Rong: Bearing fault diagnosis via kernel matrix construction based support vector machine. In: *Journal of Vibroengineering* 19 (2017), Nr. 5, pp. 3445–3461

[WDS17] Wirtz, Sebastian F. ; Demmerling, AnnaA L. ; SÖffker, Dirk: In-situ wear monitoring: An experimental investigation of acoustic emission during thread forming. In: *Structural Health Monitoring 2017* (2017), pp. 1–9

[WHW21] Wang, Yumin ; Han, Minghong ; Wu, Yaman: Semi-supervised Fault Diagnosis Model Based on Improved Fuzzy C-means Clustering and Convolutional Neural Network. In: *IOP Conference Series: Materials Science and Engineering* Bd. 1043 IOP Publishing, 2021, pp. 052043

[Wid64] Widrow, Bernard: Pattern recognition and adaptive control. In: *IEEE Transactions on Applications and Industry* 83 (1964), Nr. 74, pp. 269–277

[WJP+17] Whitford, Thomas J. ; Jack, Bradley N. ; Pearson, Daniel ; Griffiths, Oren ; Luque, David ; Harris, Anthony W. ; Spencer, Kevin M. ; Le Pelley, Mike E.: Neurophysiological evidence of efference copies to inner speech. In: *Elife* 6 (2017), pp. e28197

[WJX+21] Wang, Kaimeng ; Jing, Hongyang ; Xu, Lianyong ; Han, Yongdian ; Zhao, Lei ; Song, Kai: Cyclic response and dislocation evolution of a nickel-based superalloy under low cycle fatigue deformation. In: *Materials Science and Engineering: A* 814 (2021), pp. 141225

[WKW16] Weiss, Karl ; Khoshgoftaar, Taghi M. ; Wang, DingDing: A survey of transfer learning. In: *Journal of Big data* 3 (2016), Nr. 1, pp. 1–40

[WLGZ17] Wen, Long ; Li, Xinyu ; Gao, Liang ; Zhang, Yuyan: A new convolutional neural network-based data-driven fault diagnosis method. In: *IEEE Transactions on Industrial Electronics* 65 (2017), Nr. 7, pp. 5990–5998

[WLR+15] Wang, Xingqing ; Li, Yanfeng ; Rui, Ting ; Zhu, Huijie ; Fei, Jianchao: Bearing fault diagnosis method based on Hilbert envelope spectrum and deep belief network. In: *Journal of Vibroengineering* 17 (2015), Nr. 3, pp. 1295–1308

[WLS+19] Weichert, Dorina ; Link, Patrick ; Stoll, Anke ; Rüping, Stefan ; Ihlenfeldt, Steffen ; Wrobel, Stefan: A review of machine learning for the optimization of production processes. In: *The International Journal of Advanced Manufacturing Technology* 104 (2019), Nr. 5, pp. 1889–1902

[WLZ+16] Wang, H ; Lei, Ze ; Zhang, X ; Zhou, B ; Peng, J: Machine learning basics. In: *Deep Learn* (2016), pp. 98–164

[WMZ09] Wang, Hua ; Ma, Cuiqin ; Zhou, Lijuan: A brief review of machine learning and its application. In: *2009 international conference on information engineering and computer science* IEEE, 2009, pp. 1–4

[WP19] Wan, Shuting ; Peng, Bo: An integrated approach based on swarm decomposition, morphology envelope dispersion entropy, and random forest for multi-fault recognition of rolling bearing. In: *Entropy* 21 (2019), Nr. 4, pp. 354

[WPH17] Wickramasinghe, KC ; Perera, GIP ; Herath, HMCM: Formulation and performance evaluation of a novel coconut oil–based metalworking fluid. In: *Materials and Manufacturing Processes* 32 (2017), Nr. 9, pp. 1026–1033

[Wu17] Wu, Jianxin: Introduction to convolutional neural networks. In: *National Key Lab for Novel Software Technology. Nanjing University. China* 5 (2017), Nr. 23, pp. 495

[WWW+21] Wang, Qingfeng ; Wang, Shuai ; Wei, Bingkun ; Chen, Wenwu ; Zhang, Yufei: Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features. In: *IEEE Access* 9 (2021), pp. 45428–45440

[WZLL21] Wan, Lanjun ; Zhang, Gen ; Li, Hongyang ; Li, Changyun: A novel bearing fault diagnosis method using spark-based parallel ACO-K-Means clustering algorithm. In: *IEEE Access* 9 (2021), pp. 28753–28768

[XLX+17] Xia, Min ; Li, Teng ; Xu, Lin ; Liu, Lizhi ; De Silva, Clarence W.: Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. In: *IEEE/ASME transactions on mechatronics* 23 (2017), Nr. 1, pp. 101–110

[XT19] Xu, Fan ; Tse, Peter W.: Automatic roller bearings fault diagnosis using DSAE in deep learning and CFS algorithm. In: *Soft computing* 23 (2019), Nr. 13, pp. 5117–5128

[YFCZ18] Yi, Jing ; Fu, Sibao ; Cui, Shaohua ; Zhao, Chenglin: A deep contractive auto-encoding network for machinery fault diagnosis. In: *2018 18th International Symposium on Communications and Information Technologies (ISCIT)* IEEE, 2018, pp. 1–5

[YH17] Yang, Tianqi ; Huang, Shuangxi: Fault diagnosis based on improved deep belief network. In: *2017 5th International Conference on Enterprise Systems (ES)* IEEE, 2017, pp. 305–310

[YNB+16] Yoshimura, Natsue ; Nishimoto, Atsushi ; Belkacem, Abdelkader N. ; Shin, Duk ; Kambara, Hiroyuki ; Hanakawa, Takashi ;

Koike, Yasuharu: Decoding of covert vowel articulation using electroencephalography cortical currents. In: *Frontiers in neuroscience* 10 (2016), pp. 175

[YP19] Yadav, Om P. ; Pahuja, GL: Bearing Fault Detection Using Logarithmic Wavelet Packet Transform and Support Vector Machine. In: *International Journal of Image, Graphics & Signal Processing* 11 (2019), Nr. 5, pp. 21–33

[YPZM19] Yang, Yang ; Peng, Zhike ; Zhang, Wenming ; Meng, Guang: Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances. In: *Mechanical Systems and Signal Processing* 119 (2019), pp. 182–221

[Yu11] Yu, Jianbo: Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. In: *Mechanical Systems and Signal Processing* 25 (2011), Nr. 7, pp. 2573–2588

[YWLZ19] Yu, Hui ; Wang, Kai ; Li, Yan ; Zhao, Wu: Representation learning with class level autoencoder for intelligent fault diagnosis. In: *IEEE Signal Processing Letters* 26 (2019), Nr. 10, pp. 1476–1480

[ZCL+20] Zhang, Tianci ; Chen, Jinglong ; Li, Fudong ; Pan, Tongyang ; He, Shuilong: A small sample focused intelligent fault diagnosis scheme of machines via multimodules learning with gradient penalized generative adversarial networks. In: *IEEE Transactions on Industrial Electronics* 68 (2020), Nr. 10, pp. 10130–10141

[ZHWL14] Zhao, Peilin ; Hoi, Steven C. ; Wang, Jialei ; Li, Bin: Online transfer learning. In: *Artificial intelligence* 216 (2014), pp. 76–102

[ZKM+18] Zhang, Tianyu ; Kang, Jiawen ; Meng, Dezhuang ; Wang, Hongwei ; Mu, Zhengming ; Zhou, Meng ; Zhang, Xiaotong ; Chen, Chen: Mathematical methods and algorithms for improving near-infrared tunable diode-laser absorption spectroscopy. In: *Sensors* 18 (2018), Nr. 12, pp. 4295

[ZLC+19] Zhang, Ansi ; Li, Shaobo ; Cui, Yuxin ; Yang, Wanli ; Dong, Rongzhi ; Hu, Jianjun: Limited data rolling bearing fault diagnosis with few-shot learning. In: *IEEE Access* 7 (2019), pp. 110895–110904

[ZLLN18] Zhang, Bo ; Li, Wei ; Li, Xiao-Li ; Ng, See-Kiong: Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks. In: *Ieee Access* 6 (2018), pp. 66367–66384

[ZLT+20] Zhou, Qianwen ; Lv, Yaqiong ; Tu, Lei ; Wang, Miao ; Li, Shijie: Study of fault diagnosis for rolling bearing based on clustering algorithms. In: *2020 5th International Conference on Control and Robotics Engineering (ICCRE)* IEEE, 2020, pp. 58–62

[ZLW+20] Zhang, Zongzhen ; Li, Shunming ; Wang, Jinrui ; Xin, Yu ; An, Zenghui ; Jiang, Xingxing: Enhanced sparse filtering with strong noise adaptability and its application on rotating machinery fault diagnosis. In: *Neurocomputing* 398 (2020), pp. 31–44

[ZNZ+16] Zhang, Xin ; Ni, Xianglong ; Zhao, Jianmin ; Sun, Fucheng ; Du, Zhendong: Rolling bearing fault diagnosis using modified K-means cluster analysis. In: *Vibroengineering Procedia* 10 (2016), pp. 155–160

[ZPC17] Zheng, Jinde ; Pan, Haiyang ; Cheng, Junsheng: Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. In: *Mechanical Systems and Signal Processing* 85 (2017), pp. 746–759

[ZR15] Zhao, Shunan ; Rudzicz, Frank: Classifying phonological categories in imagined and articulated speech. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* IEEE, 2015, pp. 992–996

[ZTS+18] Zhang, Wenchang ; Tan, Chuanqi ; Sun, Fuchun ; Wu, Hang ; Zhang, Bo: A review of EEG-based brain-computer interface systems design. In: *Brain Science Advances* 4 (2018), Nr. 2, pp. 156–167

[ZYZ21] Zhi, Liu ; Yuanliang, Ju ; Zhongheng, Xie: Anti-noise motor fault diagnosis method based on decision tree and the feature extraction methods in the time domain and frequency domain. In: *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)* IEEE, 2021, pp. 71–75

[ZZH+21] Zhang, Yahui ; Zhou, Taotao ; Huang, Xufeng ; Cao, Longchao ; Zhou, Qi: Fault diagnosis of rotating machinery based on recurrent neural networks. In: *Measurement* 171 (2021), pp. 108774

[ZZL+17] Zhao, Jianmin ; Zhang, Xin ; Li, Haiping ; Ni, Xianglong ; Du, Zhendong: Health indicator selection and health assessment of rolling element bearing. In: *2017 9th International Conference on Modelling, Identification and Control (ICMIC)* IEEE, 2017, pp. 447–452

[ZZWH20] Zhang, Shen ; Zhang, Shibo ; Wang, Bingnan ; Habetler, Thomas G.: Deep learning algorithms for bearing fault diagnostics—A comprehensive review. In: *IEEE Access* 8 (2020), pp. 29857–29881

This thesis is based on the results and development steps presented in the following publications.

Journal Articles

[DWS22] Demmerling A.L., Wei X., and Söffker D.: Extended tapping torque test to differentiate metalworking fluids. Tribology International, vol. 175, 2022, pp. 107819.

[WJS22] Wei X., Jochmann F., and Söffker D.: A new approach to classify CWRU bearing states precisely, in preparation.

[WLS22] Wei X., Lee T.S., and Söffker D.: An Unsupervised Learning Approach for Metalworking Fluids Distinction Based on Acoustic Emission Signals. MDPI Sensors, 2022, submitted.

[WSS22] Wei X., Surjana A.I., and Söffker D.: Inner speech classification based on electroencephalography (EEG) signals and support vector machine (SVM). Brain Informatics, 2023, in preparation.

Conference Papers

[WLS23] Wei X., Lee T.S., and Söffker D.: A new unsupervised learning approach for CWRU bearing state distinction. European Workshop on Structural Health Monitoring, vol. 270, 2023, pp. 312-319.

[WJDS22] Wei X., Jochmann F., Demmerling A.L., and Söffker D.: Application of transfer learning in metalworking fluid distinction. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 2, 2022, pp. V002T02A003.

[WS22] Wei X., and Söffker D.: A new data processing method for Acoustic Emission signals for metalworking fluid classification. Annual Conference of the PHM Society, vol. 14, 2022, pp. 471-476.

[WDS21] Wei X., Demmerling A.L., and Söffker D.: Metalworking Fluid Classification Based on Acoustic Emission Signals and Convolutional Neural Network. European Conference of the PHM Society, vol. 6, 2022, pp. 6-6.

[WS20] Wei X., and Söffker D.: Comparison of CWRU dataset-based diagnosis approaches: review of best approaches and results. 10th European Workshop on Structural Health Monitoring (EWSHM 2020) - Special Collection, vol. 127, 2020, pp. 525-532.

In the context of research projects at the Chair of Dynamics and Control, the following student thesis have been supervised by Xiao Wei and Univ.-Prof. Dr.-Ing. Dirk Söffker. Development steps and results of the research projects and the student theses are integrated with each other and hence are also part of this thesis.

[Sur22] Surjana, A.I., Application of Machine Learning on Brain Speech Recognition, Master Thesis, September 2022

[Lee21] Lee, T.S., Establishing and application of unsupervised machine learning methods on bearing fault analysis and metalworking fluids, Master Thesis, November 2021

[Joc21] Jochmann, F., Classification of Acoustic Emission-based industry and CWRU benchmark datasets affected by variable operating conditions using Convolutional Neural Network (CNN), Master Thesis, October 2021

DOI: 10.17185/duepublico/78472
URN: urn:nbn:de:hbz:465-20230605-101214-5