Deep Learning for anomaly detection in industry 4.0

Aurélien Keleko Teguede

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Informatique

Présentée et soutenue par :

M. AURÉLIEN KELEKO TEGUEDE

le jeudi 1 décembre 2022

Titre :

Apprentissage profond pour l'aide à la détection d'anomalies dans l'industrie 4.0

Ecole doctorale :

Aéronautique-Astronautique (AA)

Unité de recherche :

Laboratoire Génie de Productions de l'ENIT (E.N.I.T-L.G.P.)

Directeur(s) de Thèse :

M. BERNARD KAMSU-FOGUEM

M. RAYMOND HOUE NGOUNA

Rapporteurs :

M. SAMIR LAMOURI, ENSAM - ARTS ET METIERS PARISTECH

Membre(s) du jury :

M. BLAISE NSOM, UNIVERSITE DE BRETAGNE OCCIDENTALE, Président

M. AMÈVI TONGNE, ECOLE NATIONALE D'INGENIEURS DE TARBES, Membre

M. BERNARD KAMSU-FOGUEM, ECOLE NATIONALE D'INGENIEURS DE TARBES, Membre

MME GISÈLE ADÉLIE MOPHOU LOUDJOM, UNIVERSITE ANTILLES GUYANE, Membre

M. RAYMOND HOUE NGOUNA, ECOLE NATIONALE D'INGENIEURS DE TARBES, Membre

Dedicate

I dedicate this humble thesis work:

To the Memory of my father Flaubert Ernot KELEKO and
my great brother Maurice Servel KELEKO who left at a very early age.

I hope that, from the world that is yours now, you will appreciate
this humble act is proof of gratitude and affection on my part.
I have always had a great thought for you,
and I pray that your soul may rest in peace.

Thanks to you, my dear father, I have followed this career.
This work results from the sacrifice you made for my education, and the person
I have become today. I will always be grateful to you for the love,
counsel, and attention you have provided me.

To you great brother, you will always remain in my thoughts and in my heart.
I also dedicate this work to you because you
were a great source of motivation.
I still remember the day I got my very first diploma in Cameroon.
Your happiness was so immense. I will always be grateful to you.

May God keep your souls in His great kingdom.

Acknowledgements

First, I would like to express my gratitude to Almighty and Merciful God for giving me the strength, perseverance, and patience to do this work.

Then, I have to express my profound gratitude to all who have supported me in this thesis. In particular, my supervisor Mr. Bernard Kamsu-Foguem, a Professor at the National Engineering School of Tarbes (ENIT) for accepting to supervise me in this project so full of great surprise and success. In fact, he guided me throughout this project by sharing with me his great ideas. I am grateful for his kindness, his constant support, his great disponibility, and the many encouragements and advice he provided me. Thanks to him, I have become more autonomous and independent throughout my research work.

I would also be grateful to Mr. Raymond Houe Ngouna, Co-supervisor of this thesis and Associate Professor at the National Engineering School of Tarbes (ENIT). This research work represents the result of more than three years of collaboration with him. Through his side, I have understood the meaning of rigor and precision in scientific research. Thanks for your patience and your advice.

I thank all the members of the jury, in particular, Mr. Blaise NSOM Professor at the University of Bretagne Occidentale (UBO), and Mr. Samir LAMOURI Professor at the Ecole Arts et Métier ParisTech who accepted to be the Referees of this doctoral thesis. I sincerely thank you for the time you have spent reading and evaluating this work. Your scrupulous proofreading, your various suggestions, feedback, and your comments have been important and constructive. Thus, we improved this manuscript thanks to your suggestions and remarks.

I would also like to thank the Examiners for their presence. In particular, I would like to thank Mrs. Gisèle Adélie MOPHOU LOUDJOM Professor at the University of Antilles for the honor she has in accepting to be the rapporteur of my thesis.

I would like to thank Mr. Amevi Tongne, Associate Professor at the National Engineering School of Tarbes (ENIT), for his help and advice in developing this research work. I also thank him for accepting to be the Examiner.

I deeply thank all my family. Especially my Mother JEANNETTE who always knew how to motivate and encourage me since my early childhood and for her many years of support and encouragement throughout my career. I hope that you will find in the accomplishment of this work the results of your efforts and the expression of my greatest gratitude.

A special thanks you to my great DESIRE for his presence and encouragement during the difficult and stressful moments. You have always been there for me when I needed it.

I thank all my other brothers and sisters for your unconditional support and affection.

I also thank MARINA who was present during his last moments.

Finally, I would like to thank my friend IURI, Edwige, and all the other people I have probably forgotten who have participated in some way in the success of this thesis project.

"Theory is when you know everything and
nothing works. Practice is when everything works,
and no one knows why.
Here we have brought together theory and practice:
Nothing works ... and no one knows why!"

Albert Einstein

Contents

List of abbreviations, and Acronyms **xvii**

List of Annotations and Symbols **xxi**

1 General Introduction **3**

1.1 Context and Motivations 3

1.2 Decision Support Tools Implementation 8

1.3 Objectives and Research Questions 10

1.4 Organization of the Manuscript 14

1.5 Publications 15

2 State of the Art of Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis **17**

2.1 Introduction 19

2.2 Contributions and Research Objectives of the Chapter 20

2.3 Overview of the Industrial Revolution and Maintenance Strategies 21

2.4 Most common AI techniques used in Predictive Maintenance 4.0 25

2.5 Methodology for the Study 30

2.6 Analysis, and Results 33

2.7 Discussion 52

2.8 Conclusion, Limitation and Future Works Orientations 54

3 Health Condition Monitoring of a Complex Hydraulic System using Deep Neural Network and DeepSHAP Explainable XAI **57**

3.1 Introduction 58

3.2 Literature Review of Condition Monitoring Applied to Hydraulic Systems . . 61

3.3 Deep Neural Network for Fault Classification and Shapley Additive exPlanations approach for Explain the Model 63

3.4 Developed Framework 75

3.5 Hydraulic System and Sensor Data Description 79

3.6 Results of Developed Framework 81

3.7 Discussion 92

3.8 Conclusions and Future Work 98

4 Physical-Informed Neural Networks and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding (FSW) 101

4.1 Introduction 103

4.2 Overview of the Process Modeling Techniques 106

4.3 Thermomechanical Process: Friction Stir Welding (FSW) 110

4.4 Theories of the Numerical Simulation of the Thermomechanical Problem . . . 111

4.5 Proposed Methodology 115

4.6 Results and Discussions 127

4.7 Conclusion 133

5 General Conclusions and Futures works 135

5.1 Summary of Contributions 136

5.2 Limitations of the Proposed Approaches 139

5.3 Suggestions for Future Works 140

A Appendix - Some results of PINN Module 141

Bibliographie 172

List of Figures

1.1 Main industrial revolutions [1] 4

1.2 The main pillars or keys components that characterize the fourth Industrial Revolution [5] 6

1.3 Maintenance strategies in the industry. 8

1.4 Main steps in the implementation of decision-making tools 9

1.5 General diagram of the organization of the manuscript 15

2.1 Historical perspective of Industrial Revolutions and their associated inspection or monitoring techniques. 23

2.2 Different types of techniques or approaches for monitoring or maintenance in the context of Industry 4.0. (Adapted to [55]). 24

2.3 Methodology framework for bibliometric analysis. Each color corresponds to a step of methodology. The different steps represent the methods or strategies used to perform this study 31

2.4 Pie Chart: The graphic represents the types of retrieved documents over the last 20 years 33

2.5 Annual scientific production published in WoS journals over the last 20 years. 35

2.6 Network visualization for most productive journal. 36

2.7 Network visualization for Publication highly Co-authorship. Each cluster is represented by a color. To interpret the results, and the color of the legends in this figure, the reader can refer to the Web version of this manuscript. . . . 39

2.8 Network visualization for international collaboration affiliation. Each group is represented by a color. 39

2.9 World map of the country-level scientific productivity, for documents collected on WoS over the last 20 years. The color scale is given by the number of articles, dark blue: high productivity, light blue: low productivity. 42

2.10 Network visualization for international collaboration. 43

2.11 Network visualization for Co-citations. Cluster 1: yellow color, cluster 2: red color, Cluster 3: green color, and Cluster 4: blue color. 44

2.12 Density visualization map of the most frequently related terms in retrieved articles on WoS. The frequent terms were visualized using VOSviewer. 46

2.13 Conceptual Map, and keywords clusters (minimum number of documents is 5, and method is MCA. Each cluster is represented by a distinct color (Cluster 1: red color, and cluster 2: blue color). 47

2.14 Word dynamics for the keyword plus 48

2.15 Topic trend analysis of the collected documents over the last 20 years. 48

2.16 Strategic diagram (adapted from [137]) 49

2.17 Strategic map of the author's keyword. To interpret the results, and the color of the legends keyword in this figure, the reader is referred to the color. . . . 50

3.1 Architecture of the fully connected Deep Neural Network Multi-class classification model with four hidden layers, each layer contains nine neurons. 66

3.2 Framework of the "Black-Box" AI compared with the explainable XAI approaches. 68

3.3 The interests of all stakeholders in the use and importance of the XAI approaches. 69

3.4 Diagram of the proposed framework. The workflow has two principal modules: (a) The DNN multi-class classification module aims to predict the different conditions for failure of the hydraulic system components. (b) The explainable DeepSHAP module provides insight into the decision-making process of the classifier model. 78

3.5 Diagram of the hydraulic test bench: The test rig is composed of two main circuits, namely the primary working circuit to control the load and the secondary cooling-filtration circuit, which are connected by the oil tank. An upper circuit is the primary working circuit, it is connected to the lower circuit which provides cooling and filtration via the oil tank; the sensors are highlighted in blue [198], [236]. 80

3.6 Frequency distribution of the sensors data: The histogram represents an average of each data cycle. 83

3.7 Confusion Matrices for the DNN Multi-classification module. Each matrix represents the frequency of the misclassification and their relative types of errors. Sub-figures (a), (b), (c), and (d) represent respectively the confusion matrix of the Cooler Condition, Internal Pump Leak, Hydraulic Accumulator, Valve Condition, and Stable Flag 87

3.8 Force Plot and local interpretation for the valve conditions: This plot describes the function's output using the sum of these effects. Furthermore, the importance of the force of each feature is explained at different moments during the model training and decision-making. Thus, the features with a positive impact (contributing to the prediction being higher than the baseline value) are highlighted in red. In contrast, the features with a negative effect (contributing to the prediction being lower than the baseline value) are in blue. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN. 90

3.9 Force Plot describes the local interpretation) for the cooler condition classification state obtained from the DeepSHAP method module. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN. 91

3.10 Force Plot shows the local interpretation for the internal pump leakage classification state obtained from the DeepSHAP method. In particular, sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN. 91

3.11 Force Plot is the local interpretation for the hydraulic accumulator classification state obtained from the DeepSHAP approach. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN. 92

3.12 Force Plot represents the local interpretation for the stable flag classification state obtained from the DeepSHAP method. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN. 92

3.13 Summary Plot (global interpretation) for the Cooler condition classification. The high Shapley values are in red dots, and the lower ones are in blue. The figures show the global importance of the variables. In Particular, sub-figure (a) illustrates the absolute average of the Shapley values, and sub-figure (b) shows the Shapley values of each feature according to their order of importance. 93

3.14 Summary Bar Plot represents the global importance of each component of the hydraulic conditions. In particular, sub-figures (a), (b), (c), and (d) show respectively the global importance feature or contribution of each sensor in the decision-making models for the components of the hydraulic system (Valve conditions, Hydraulic accumulator, Internal pump leak and Stable flag) . . . 94

3.15 Summary plot: Sub-figures (a), (b), (c) and (d) show respectively the global importance or contribution for each sensor/feature contribution to the decision-making DNN model. 95

4.1 Illustration of the number of possible ways in which AI can be performed [246]. 104

4.2 Schematic representation of the approaches used to address predictive tasks on a complex system. 107

4.3 Scheme of the main phases of the thermomechanical Friction Stir Welding. . 110

4.4 Model geometry of the thermomechanical Friction Stir Welding 112

4.5 Model meshing and boundary conditions 112

4.6 Diagram of the Friction Stir Welding process: The blue arrows represent the rotation velocity imposed on the pin, while the advancing velocity is replaced by a velocity imposed on the plates in the opposite direction (red arrows) [308]. 113

4.7 Detailed flow chart of the developed methodology. This methodology is mainly based on two modules, namely the numerical simulation module and the PINN approach applied to 2D data. 117

4.8 Formal neuron represented by x_n inputs, output y and a given activation function. 118

4.9 Dropout regularization method applied to the weight of the Neural Network. Figure (a) shows the initial state of the network, and figure (b) indicates the momentary suppression of the neurons during the network training. 122

4.10 Schematic of physics-informed neural network (PINN). This schematic represents a detailed description of block five of the figure 4.7). PINNs are composed of residues (losses) of differential equations 4.21. The FCNN takes as inputs the vector $(X, t) = [(x, y), t]$ where (x, y) and t are respectively the spacial coordinates and times. This network is trained to approximate the physics solutions $\hat{y} = [\hat{u}, \hat{v}, \hat{p}]$. We used Automatic Differentiation (AD) to compute the derivatives of parameters with respect to the input vector. Then AD helps to express the residuals of the governing equations in the combined loss function. This function is generally composed of several terms weighted by different coefficients. The last step containing all residuals is known as the feedback mechanism. This aims to minimize the loss function, using the Adam optimizer, based on some learning rate in order to obtain the parameters θ of the network. In fact, the parameter θ of the FCNN and the parameter λ of the unknown PDE can be estimated simultaneously by minimizing the combined loss function. . . . 125

4.11 Simulation results of the $2D$ process (Contour plots): Contour plots represent the evolution of the velocities and pressure subsequently at the beginning, middle, and end of the process. For each step, the time steps are respectively equal to 0.0060, 0.0765, and 0.150 sec). 127

4.12 Losses functions: Each figure (a), (b), (c) and (d) respectively represents the loss function related to the variable u, v, p and the derivatives for the NSE. Furthermore, sub-figure (e) show the combined loss function defined by the equation 4.21, with respect to the number of training epochs 129

4.13 A comparison of the results of the $2D$ data at beginning of the process. Sub-figures (a) and (b) represent the realistic simulation of the contour plot. Then, sub-figures (c) and (d) are the predicted values with less than 20% of the training set. Finally, sub-figures (e) and (f) show the best result obtained with more than 70% of the training set. 130

4.14 A comparison of the results of the $2D$ data at the middle and end of the process. The first two contours plots (a) and (b) represent the data realistic simulation and the last two contours plot (c) and (d) represent the best result obtained . 131

A.1 Results of realistic data (simulation of the $2D$ FSW process). Contour plot represent the evolution of the velocities and pressure subsequently at the beginning. The time steps are respectively equal to 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec). 143

A.2 Results of realistic data (simulation of the $2D$ FSW process). Contour plot represent the evolution of the velocities and pressure subsequently at the middle of the process. The time steps are respectively equal to 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec). 144

A.3 Predicted values at the beginning of the process with less than 20% of the training set. The time steps are equal o 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec. 145

A.4 Predicted values at the beginning of the process with less than 20% of the training set. The time steps are equal 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec). 146

A.5 The best result obtained with more than 70% of the training set. Predicted values at the beging of the process. The time steps are equal to 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec 147

A.6 The best result obtained with more than 70% of the training set. Predicted values at the end of the process. The time steps are equal to 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec 148

List of Tables

2.1 Modeling approaches for fault detection, and diagnosis in predictive maintenance. . . . 26

2.2 Main information and statistics regarding the collection published between 2000, and 2021 on WoS. . . . 34

2.3 Productivity: Annual number of published articles between 2000-2021 on WoS. ND is number of the documents, and ND (%) is a number of documents in percent. . . . 36

2.4 Most productive journal sorted by the publication number (NP), most impact's document (h-index), and most cited journals (TC). . . . 37

2.5 Most cited authors: Authors are ordered by a Total of Citations (TC) index . 38

2.6 Most relevant affiliations ordered by a number of articles. . . . 40

2.7 The most cited organizations ordered by TC index. . . . 41

2.8 Most productive and cited countries ordered by the frequency publication, or productivity (years: 2000-2021). . . . 42

2.9 Most globally cited scientific publications. . . . 43

2.10 Top 20 of the most frequent authors' keywords. . . . 45

2.11 Cluster of co-occurrence network author's keyword . . . 45

2.12 Strategic Map author's keywords. Each cluster is represented by the main theme and its positioning in relation to the current literature. . . . 50

3.1 Summary of work related to monitoring conditions applied to sensor data collected from the hydraulic system. . . . 63

3.2 Physical data (quantitative data) collected by the sensor. . . . 81

3.3 Monitored parameters: Categorical data of the hydraulic test bench. Each variable represents a system operating multi-state. . . . 82

3.4 Table shows the performance of the DNN classifier model for Cooler conditions. We use several values of learning rate to train the DNN. The best model selection is chosen according to the highest Accuracy value. . . . 85

3.5 Summary and parameters numbers of the best DNN classifier model. . . . 86

3.6 Performance result of the proposed DNN multi-class classification model . . . 86

3.7 DNN classifier model results for the multi-class classification of degradation levels of each state of the hydraulic system. 88

3.8 DNN classifier model results for the multi-class classification of degradation levels of each state of the hydraulic system. 89

4.1 Some applications of PINN models and their extensions. 109

4.2 Mechanical properties of material 115

4.3 Variations of the hyper-parameters of the PINN model 128

List of Abbreviations, and Acronyms

Adam	*Adaptive Moment Estimation*
AI	*Artificial Intelligence*
ALA	*Adaptive Linear Approximation*
ANN	*Artificial Neural Networks*
AMD	*Adaptive Moment Estimation*
AE	*Auto Encoder*
AI	*Artificial Intelligence*
ALE	*Arbitrary Lagrangian-Eulerian*
CAFE	*Cellular Automata Finite Element*
CE	*Cross-Entropy*
CdM	*Condition Maintenance*
CM	*Condition monitoring*
CBM	*Condition-Based Maintenance*
cGAN	*Conditional Generative Antagonistic*
CFD	*Computational Fluid Dynamic*
CPU	*Central Processing Unit*
CRFS	*Correlation and Redundancy-aware Feature Selection*
CNN	*Convolution Neutral Networks Networks*
DT	*Decision Tree*
DeepLIFT	*Deep Learning Important FeaTures*
DeepSHAP	*Deep SHapley Additive exPlanations*
DL	*Deep Learning*
DOI	*Digital Object Identifier*
DNN	*Deep Neural Network*

DT	*Decision Tree*
EMD	*Empirical Mode Decomposition*
FP	*False Positives*
FN	*False Negatives*
FEM	*Finite Element Method*
FVM	*Finite Volume Method*
FDM	*Finite Difference Method*
FCNN	*Fully-Connected Neural Network*
FCN	*Fully Convolutional Networks*
FSW	*Friction Stir Welding process*
FPINN	*Fractional Friction Stir Welding process*
FVM	*Finite Volume Method*
GPU	*Graphics Processing Unit*
GBoost	*Extreme Gradient Boosting*
GAN	*Generative Adversarial Networks*
GPU	*Graphical Processing Unit*
I4.0	*Industry 4.0*
IoT	*Internet of Things*
IEEE	*Institute of Electrical and Electronics Engineers*
ISSN	*International Standard Serial Number*
KL-LIME	*Kullback-Leibler LIME*
K-NN	*K-Nearest Neighbors*
LDA	*Linear Discriminant Analysis*
LR	*Learning Rate*
LSTM	*Long Short-Term Memory*
LIME	*Local Interpretable Model-Agnostic Explanations*
LRP	*Relevance Back-Propagation*
MCA	*Multiple Correspondence Analysis*

MCC *Mathews Correlation Coefficient*

MPS-LIME *Modified Perturbed Sampling for LIME*

MSE *Mean Squared Errors*

ML *Machine Learning*

MIT *Massachusetts Institute of Technology*

NB *Naïve Bayes*

NP *Number of Publications*

ND *Number of Documents*

NSE *Navier-Stokes Equation*

NN *Neural Network*

PLR *Penalized Linear Regression*

PAS *Predictive Analytics System*

PCA *Pincipal Component Analysis*

PdM *Predictive Maintenance*

PDEs *Partial Differential Equations*

QLIME *Quadratic LIME*

QDA *Quadratic Discriminant Analysis*

RQ *Research Question*

RMSE *Root Mean Squared Errors*

ReLU *Rectified Linear Unit*

RMSE *Root Mean Square Error*

RQ *Research Question*

RUL *Remaining Useful Life*

RNN *Recurrent Neural Networks*

RCNN *Regions with CNN*

RBM *Restricted Boltzmann Machine*

RF *Random Forest*

RBM *Restricted Boltzmann Machine*

SMV *Saliency Map Visualization*

SVM *Support Vector Machine*

SLIME *Sound-LIME*

SHAP *SHapley Additive exPlanations*

SGD *Stochastic Gradient Descent*

SPH *Smoothed Particle Hydrodynamics*

TN *True Negatives*

TC *Total of Citations*

TL *Transfer Learning*

TP *True Positives*

URL *Uniform Resource Locator*

USA *United States of America*

VGG *Visual Geometry Group*

XAI *eXplainable Artificial Intelligence*

List of Annotations and Symbols

X	*Vector of input data*
x_i	*Set of all input features*
x_0	*A given point on the images x*
X_{max}	*Maximum value of the vector X*
X_{min}	*Minimum value of the vector X*
X_{Norm}	*Min-Max Normalization of the vector X*
x^*	*Input that maximizes the activation of a given hidden unit*
Y	*Vector of output data*
y_i	*Set of all output features*
b_i	*Bias coefficient of the neural network model*
$f(.)$	*Neural network model*
h_j	*Activation function of an output layer j*
C	*Number of classes*
t_i	*Class label of output Y*
O	*x intput of the neurons*
r	*Reference input*
$R(z_j)$	*Relevance score of activation output z_j*
$R(x)$	*Relevance score of individual input x*
$S_c(.)$	*Score function of the class c*
w_{ij}	*Kernel coefficient*
G	*Family models*
g	*Class of interpretable G models*
$ReLU(h_j)$	*ReLu activation function of hidden layer h_j*
$z^{'}$	*Coalition vector for SHAP*
M	*Maximum coalition size*

$g(z^{'})$	*Contributions of the biases and the individual characteristics or the value predicted by the model for this instance*
I	*Input image*
V_j	*Shapley mean absolute value for feature j*
F	*Set of all features*
S	*Subset of all features*

Abstract — Industry 4.0 (I4.0) corresponds to a new way of planning, organizing, and optimizing production. Therefore, the increasing exploitation of production systems through the presence of many Internets of Things (IoT) devices, and digital transformation offers new opportunities to make factories intelligent and do smart manufacturing. However, there are many challenges in realizing the potential of these new technologies. One approach to addressing these challenges is introducing more automation throughout the production process. This increases the availability, profitability, effectiveness, and environmental responsibility or sustainability of the plant. This thesis focuses on factory automation via the development of decision-making tools based on data-driven and physics AI models. Besides, the theoretical aspects, the contribution, and the originality of our study consist in developing hybrid, explainable and generalizable models for Predictive Maintenance (PdM) by using Deep Learning (DL) coupled with explainable techniques. Thus, we have developed two approaches to explaining the DL model: By extracting the local and global knowledge from the learning processes to support more transparent decision-making rules through Explainable Artificial Intelligence (XAI) and by introducing knowledge or physical laws to inform and guide the DL model. For this purpose, our research will focus on three main points:

Firstly, we will provide a state of the art of anomalies detection and PdM 4.0 approaches in I4.0. Thus, we will exploit an advanced bibliometric analysis to retrieve and analyze relevant documents from the scientific database Web of Science (WoS).This analysis gives us some useful guidelines to help researchers and practitioners to understand the main challenges and the most relevant scientific issues related to AI and PdM. Secondly, we have developed two frameworks, based on Deep Neural Networks (DNNs). The first framework is formed by two modules such as DNN and Deep SHapley Additive exPlanations (DeepSHAP). DNN module consists to address the unbalanced multi-class classification tasks applied to the hydraulic system conditions. Despite their performance, some questions arise about the reliability of DNN as a "black-box" model for decision-making and the possible ethical, impacts on stakeholders. To address these issues, a second module based on DeepSHAP is being developed for the model's explainability. DeepSHAP shows the importance and contribution of each feature in the decision-making by the models. In addition, it promotes the understanding of the process and guides humans to better understand, interpret, and trusts the AI models.

The second hybrid framework is known as Physical-Informed Deep Neural Networks (PINN) for regression tasks. This aims to predict the states of Friction Stir Welding. PINN consists of introducing the explicit knowledge or physical constraint into the learning algorithm. This provides better knowledge and forces the model to follow the process topology. Once trained, the PINN can substitute the numerical simulation of the FSW process which is computationally time-consuming. In summary, this work opens new and promising perspectives on the explainability of AI models applied to PdM 4.0. In particular, the exploitation of these frameworks contributes to more accurate knowledge about of the investigated system.

Keywords: *Industry 4.0 (I4.0), Predictive Maintenance (PdM), Friction Stir Welding (FSW), Anomaly detection, Physical Informed Neural Networks (PINN), Trustful AI, eXplainable Artificial Intelligence (XAI)*

Résumé— L'industrie 4.0 (I4.0) correspond à une nouvelle façon de planifier, d'organiser, et d'optimiser les systèmes de production. Par conséquent, l'exploitation croissante de ces systèmes grâce à la présence de nombreux objets connectés, et la transformation digitale offrent de nouvelles opportunités pour rendre les usines intelligentes et faire du smart manufacturing. Cependant, ces technologies se heurtent à de nombre défis. Une façon de leurs d'appréhender consiste à automatiser les processus. Cela permet d'augmenter la disponibilité, la rentabilité, l'efficacité et de l'usine. Cette thèse porte donc sur l'automatisation de l'I4.0 via le développement des outils d'aide à la décision basés sur des modèles d'IA guidés par les données et par la physique. Au-delà des aspects théoriques, la contribution et l'originalité de notre étude consistent à implémenter des modèles hybrides, explicable et généralisables pour la Maintenance Prédictive (PdM). Pour ce motif, nous avons développé deux approches pour expliquer les modèles: En extrayant les connaissances locales et globales des processus d'apprentissage pour mettre en lumière les règles de prise de décision via la technique l'intelligence artificielle explicable (XAI) et en introduisant des connaissances ou des lois physiques pour informer ou guider le modèle. À cette fin, notre étude se concentrera sur trois principaux points :

Premièrement, nous présenterons un état de l'art des techniques de détection d'anomalies et de PdM4.0. Nous exploiterons l'analyse bibliométrique pour extraire et analyser des informations pertinentes provenant de la base de données Web of Science. Ces analyses fournissent des lignes directrices utiles pouvant aider les chercheurs et les praticiens à comprendre les principaux défis et les questions scientifiques les plus pertinentes liées à l'IA et la PdM. Deuxièmement, nous avons développé deux frameworks qui sont basés sur des réseaux de neurones profonds (DNN). Le premier est formé de deux modules à savoir un DNN et un Deep SHapley Additive exPlanations (DeepSHAP). Le module DNN est utilisé pour ressoudre les tâches de classification multi-classes déséquilibrées des états du système hydraulique. Malgré leurs performances, certaines questions subsistent quant à la fiabilité et la transparence des DNNs en tant que modèle à "boîte noire". Pour répondre à cette question, nous avons développé un second module nommé DeepSHAP. Ce dernier montrant l'importance et la contribution de chaque variable dans la prise de décision de l'algorithme. En outre, elle favorise la compréhension du processus et guide les humains à mieux comprendre, interpréter et faire confiance aux modèles d'IA. Le deuxième framework hybride est connu sous le nom de Physical-Informed Deep Neural Networks (PINN). Ce modèle est utilisé pour prédire les états du processus de soudage par friction malaxage. Le PINN consiste à introduire des connaissances explicites ou des contraintes physiques dans l'algorithme d'apprentissage. Cette contrainte fournit une meilleure connaissance et oblige le modèle à suivre la topologie du processus. Une fois formés, les PINNs peuvent remplacer les simulations numériques qui demandent beaucoup de temps de calcul. En résumé, ce travail ouvre des perspectives nouvelles et prometteuses domaine de l'explicabilité des modèles d'AI appliqués aux problématiques de PdM 4.0. En particulier, l'exploitation de ces framework contribuent à une connaissance plus précise du système.

Mots clés: *Industrie 4.0, Maintenance Predictive, Soudage par r friction-malaxage, Détection d'anomalies, Réseaux de neurones informés par la physique, IA fiable, Intelligence Artificielle Explicable*

Chapter 1

General Introduction

Sommaire

1.1 Context and Motivations 3
1.1.1 Industrial Revolution 3
1.1.2 Industry 4.0 and Keys Technologies 4
1.1.3 Maintenance Strategies 7
1.1.4 Predictive Maintenance (PdM) 7
1.2 Decision Support Tools Implementation 8
1.3 Objectives and Research Questions 10
1.3.1 Objectives 10
1.3.2 Research Questions 10
1.3.3 Collection of Data 12
1.3.4 Research Scope and Contributions 13
1.4 Organization of the Manuscript 14
1.5 Publications 15

1.1 Context and Motivations

1.1.1 Industrial Revolution

At the beginning of the 2000s, the manufacturing industry underwent a major change due to new market expectations and the introduction of various technologies to control production activities. Prior to any development, we thought it appropriate to give a brief reminder of the main revolutions that this industry has undergone (Figure 1.1). The first industrial revolution (Industry 1.0) occurred with the development of mechanics, the exploitation of coal, and the introduction of the steam engine. The second industrial revolution or Industry 2.0 has introduced electricity, transport development, and mass production at a reduced cost. The third revolution (Industry 3.0) is identified through the exploitation of new information technologies, such as electronics and telecommunications. Finally, the fourth revolution (Industry 4.0) was presented for the first time at the Hannover Fair in 2013. In the next subsection, we will give more details including the highlights and challenges of this last industrial revolution.

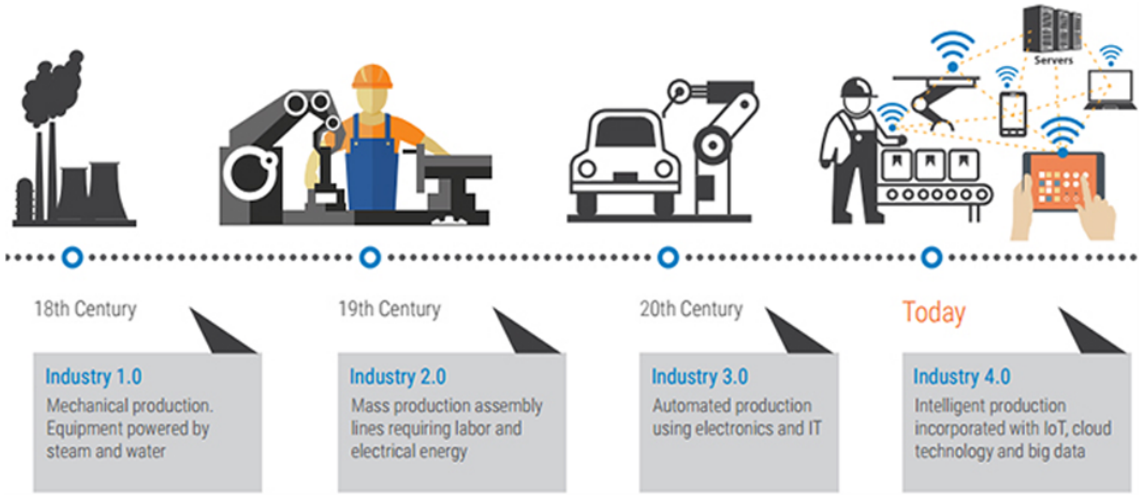


Figure 1.1: Main industrial revolutions [1]

1.1.2 Industry 4.0 and Keys Technologies

In recent years, Industry 4.0 (I4.0) has contributed to addressing new challenges in the organization of customized production means. Industrialists have to deal with increasing worldwide market competition and demands on different strategies. The definition of I4.0 depends on the application context and the research domain [2]. The German digital association Bitkom reports that there are more than 100 relevant explanations for the term Industry 4.0 [3]. Although there is strong agreement on the principles, providing a uniform definition of I4.0 is not easy. for example, articles [4], [5] describe Industry 4.0 as a combination of terms and interconnected digital technologies. We have given our own definition of this context. Thus, I4.0 can be considered as the combination of some technologies that contribute to developing, automating decision tools, and exchanging real-time data in the process. In other words, Factory 4.0 can be seen as a dynamic and integrated system that controls all the value chains of a product's life cycle. This can be identified as a new way of scheduling and organizing resources. I4.0 has aroused remarkable interest among stakeholders, including researchers, due to the emergence of new technological advances (e.g., Big Data, massive data collection, and application of advanced data science techniques). In addition, the rise in processing power, the exploitation of AI models for monitoring systems, assisted PdM, and the widespread use of distributed control systems have contributed significantly to the spread and development of I4.0. However, the industry must respond to many challenges; it should be environmentally responsible, and economical in energy and raw materials [6]. To make the factory more competitive and sustainable, financial strategies (optimizing production costs and increasing financial gains) have to be instituted [7]. Moreover, the factory must respect social, and ethical constraints, and political regulations [8]. To address these new challenges, it is necessary to digitalize the production plant by integrating automation into the entire manufacturing process [1], [9], and [10]. This automation can be achieved through cyber-physical systems, communicating sensors, and intelligent and autonomous robots. These technologies

contribute to increasing the productivity, availability, quality, and innovation of the products [11]. However, the opportunities provided by these new technologies raise a number of challenges. Figure 1.2 shows the different key components or technological advances which characterize the I4.0.

(a) Big Data and Analytics: The increasing volume of the data generated by numerous heterogeneous and different sources (interconnected machines or pieces of equipment, and production systems) are at the essence of the newly emerging research issues in I4.0. Therefore, a detailed analysis of the given data using data mining techniques is used to deploy the decision-making support tools. The benefits of these tools are numerous, such as optimizing manufacturing reducing the number of product failures, reducing production line downtime, and increasing the life of the equipment.

(b) Autonomous Robots: In the manufacturing industry, robots can interact or collaborate with each other (Machine-Machine) and with humans (Human-Machine). The exponential exploitation of these interconnected, autonomous, flexible, and cooperative robots has become a requirement. Robots contribute significantly to the production process and can be more efficient and autonomous than humans. For example, these robots help in the execution of complex and repetitive tasks by adapting to several situations (e.g., assembling and packaging products). In addition, its have the ability to automatically learn certain actions or tasks (e.g., part recognition).

(c) Simulation: It is used intensively in a wide spectrum of industrial operations. For example, it is possible to operate virtual machines which can simulate or reproduce the physical world in real-time (e.g., machining a part). The virtual machine exploits the data from the physical machines. Industrialists can increase performance by implementing this technique, the costs and production times are reduced, and the processes are optimized.

(d) Horizontal and Vertical System Integration: This technology facilitates the integration of the several components that constitute the ecosystem of I4.0, including the automation of processes. In addition, System Integration contributes to more coherence and collaboration between all the actors (e.g. customers and suppliers), functions (e.g. workers and production managers), production processes and products (raw materials, semi-finished and final products), and autonomous robots.

(e) Industrial Internet of Things: I4.0 increasingly exploits embedded systems and interconnected technologies such as IoT, sensors, databases, big data infrastructures, robots, and machines. This promotes communication and interaction between the devices. In addition, analyses and decision-making frameworks are based on the data collected through the IoT or sensors.

(f) Cybersecurity: The deployment of interconnected technologies can be subject to multiple cyber risks like spamming or malware. Consequently, it is advisable to secure critical industrial systems and production lines against these attacks. In order to address these vulnerabilities, it is indispensable to equip the industry with sophisticated and advanced systems having secure and reliable communication protocols.

(g) Cloud: The industry is increasingly using cloud-based applications to share information or data between various cloud-based software and other entities in the companies. The performance of clouds is advancing, with their connection capabilities and their responsiveness being below a few milliseconds. This advanced technology facilitates the deployment of systems for the collection of data. We can note that these data contain the machine settings and the systems which monitor and control the processes.

(h) Additive Manufacturing: Industrialists use additive manufacturing (for example 3-D printing) to improve the production processes and reduce the costs of fabrication and design of pieces (light and complex shapes). In addition, it offers advantages for the construction of prototypes in small batches and customized designs.

(i) Augmented-reality-based: For the purpose of decision-making, systems based on augmented reality have an important function in the optimization of the industry. In addition, its operation provides real-time information that will help users to take appropriate actions to ensure continuous production. Furthermore, they can receive feedback on the specific configurations, operational parameters, and maintenance instructions that they have to monitor to solve a given task.

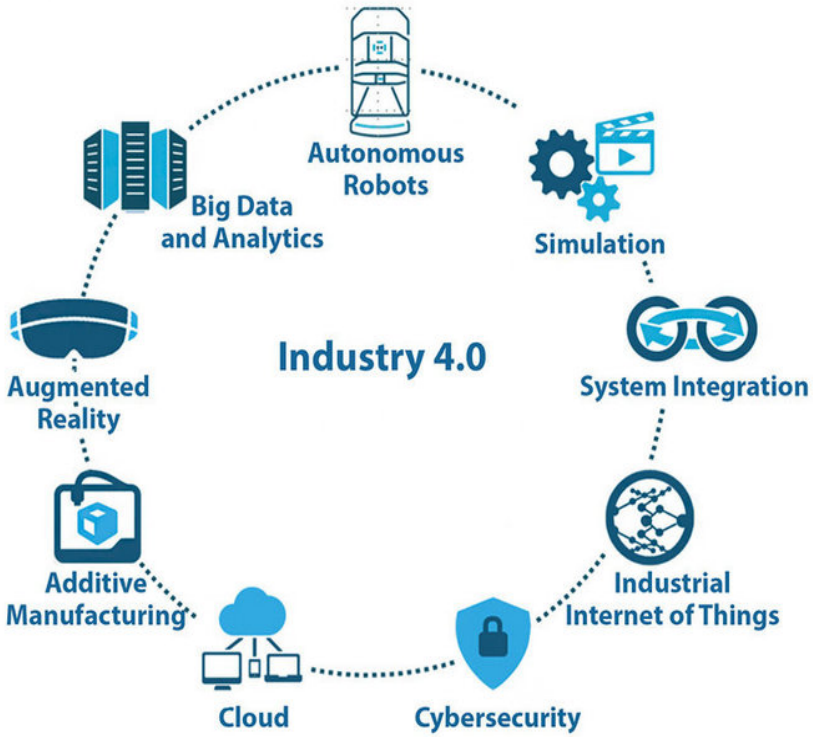


Figure 1.2: The main pillars or keys components that characterize the fourth Industrial Revolution [5]

In order to ensure interaction or communication in real-time, industrial companies use these components to connect the entire production system. This builds a new virtual world

from which we can drive the physical world. There have been several studies to highlight the different aspects, key points, and importance of these pillars in I4.0 [4], [12], [13]. We focus mainly on the data analysis and simulation components. These components are necessary to develop decision-making tools for industrial maintenance strategies such as PdM4.0.

1.1.3 Maintenance Strategies

As outlined in the first section, I4.0 must respond to new issues and opportunities such as profits, profit maximization, mass, and specialized production. These challenges are driven by technical imperatives (e.g., minimizing machine downtime and maximizing component life). These technical factors have an impact on economic questions such as the reduction of production and maintenance costs. In addition, the quality and reliability of products, and the safety of assets and services are key requirements for I4.0. In order to meet these new requirements, we use information or data that characterizes the machines/equipment. Very often, this information describes failures, breakdowns, or anomalies that occur randomly over time. If countermeasures are not anticipated, the production lines may not work or may work abnormally. To anticipate these possible failures, we can implement innovative and sophisticated diagnostic strategies and tools in industrial maintenance operations. Industrial maintenance is considered a combination of actions and management techniques that can be applied to ensure the performance of equipment over time. It is important to note that these strategies are very often periodic and do not always depend on the state of the equipment. Maintenance strategies can be classified into several categories (see figure 1.3).

- Corrective maintenance consists in performing actions when a machine has some defects or breaks down. In this case, the equipment and production line impacted will cease to operate until the failure is repaired.
- Preventive maintenance aims to reduce the probability of failure of industrial components or devices. It is performed at specific frequencies or periods. However, it does not guarantee the continuous functioning of the equipment.
- We will discuss the predictive maintenance approach, in particular on Condition Monitoring (CM) in the next section. PdM is a technique that supports the optimal functioning of all equipment and machines by eliminating or reducing the occurrence of breakdowns and optimizing the planning of maintenance work according to the technical situation. The CM approach exploits sensor data to monitor the condition of the equipment over time while in operation.

1.1.4 Predictive Maintenance (PdM)

In industrial processes, productivity decreases are often due to anomalies associated with equipment degradations, mainly when identified at an early stage. Thus, manufacturers are

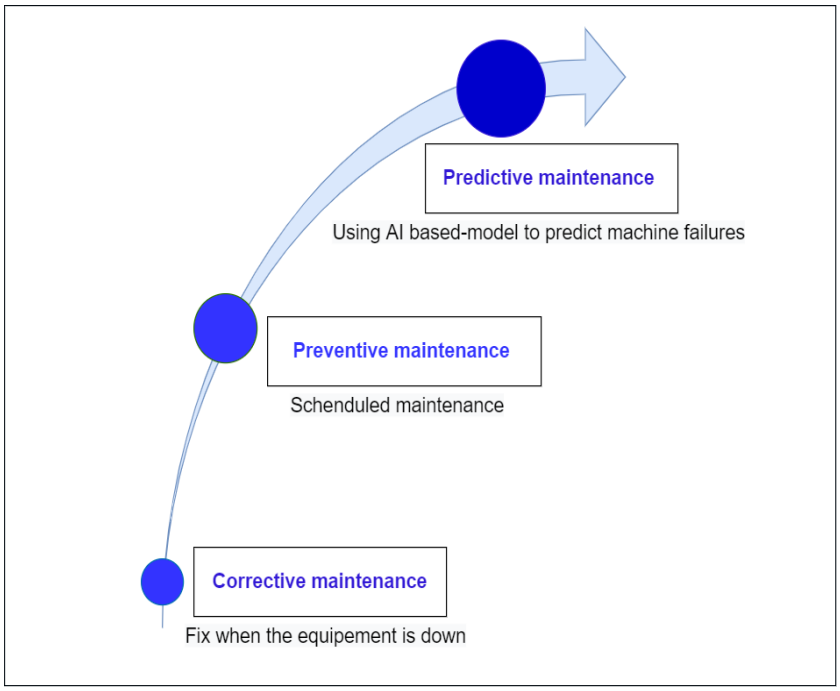


Figure 1.3: Maintenance strategies in the industry.

increasingly adopting automated predictive maintenance in their operations and processes. PdM4.0 is driven by IoTs and AI frameworks, thus, it uses information from the production process to train the algorithm able to optimize flows or supply chains. In addition, its algorithms can detect a failure at an early stage and propose appropriate actions or countermeasures to be implemented to ensure the continuous operation of the process. In addition, PdM can adapt to maintenance routines and user needs. In particular, it can perform computer-aided design instructions without additional system programming. PdM 4.0 can also integrate modules that allow monitoring of its own condition, decision-making process components, sensors, and machines. In addition, it is necessary to ensure that the tools do not interfere with the proper functioning of other machines or components.

1.2 Decision Support Tools Implementation

Figure 1.4 shows the process of implementing PdM technologies which are composed of several steps : (a) Collect heterogeneous and massive data, and store them in an accessible and secure database system. (b) Identify critical assets via formalized techniques for creative and collective problem solving such as brainstorming. (c) Explore and analyze historical data by using advanced data mining techniques. (d) Develop AI-based models to predict the failures of critical systems. (e) Combine business knowledge and prediction results for optimal decision-making. (d) Deploy and validate the tool on the mentioned systems. The resulting models provide information to anticipate failure points and possible mechanical failures. They thus

facilitate the decision-making process for maintenance activities in order to avoid downtime. In this context, the industry can be transformed into a predictive industry [14]. In addition, these innovative technologies combined with machine Condition Monitoring systems offer new possibilities for management [15], control, efficiency improvement and reliability of industrial systems [16].

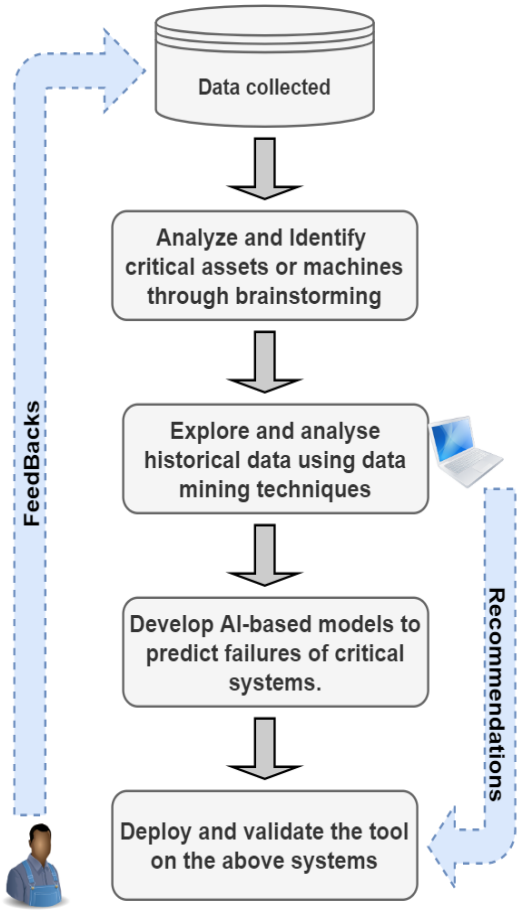


Figure 1.4: Main steps in the implementation of decision-making tools

The decision-making tool is based on a Deep Learning (DL) known as a "black-box" model having the internal structure unknown. Furthermore, the decision rules and the choice of feature variables that participate in the models are also unknown. In some cases, the collected data are unrepresentative; these anomalies involve data that are incorrectly or partially labeled and contain missing values and noise. To address the issues of data quality and quantity, several approaches have been proposed: Transfer learning, numerical simulation, generate fake data by AI approaches. Furthermore, to predict the failures of a system or a component, we can use different approaches that could depend on the nature of the data or problem: knowledge-based modeling [17]–[19], system physics-based modeling [20]–[22], data-based modeling [23]–[25] and the hybrid approach [26]–[28]. In this research, we will focus on the last two approaches. In particular, we will exploit Deep Neural Networks (DNN) to solve supervised learning tasks such as classification and regression.

DNNs are "black-box" models, involving a combination of powerful learning algorithms with many hyper-parameters and layers of varying depth. Despite their many applications and performances, the lack of transparency and explanation of the decision-making rules can be a major obstacle to their exploitation. In addition, there are questions about the large-scale application of tools based on AI techniques and their ethical impact, transparency, trust, fairness, or privacy rules. In addition, practitioners are interested in the explanation paradigm of complex AI algorithms. The integration of explanatory approaches is important for the understanding of "black-box" models. The main objective of the explanatory paradigm of algorithms is to provide answers to the different expectations, interests, goals, and needs of all stakeholders, including citizens, regulators, governments, and experts in the domain.

1.3 Objectives and Research Questions

1.3.1 Objectives

PdM4.0 offers new opportunities for the optimization of production chains. In this thesis, we focus on the modeling of two complex processes: the FSW process and the hydraulic system. The main objective of our research work is to develop decision-making tools based on DL models. To satisfy performance, reliability, and confidence criteria, we will show that the developed models are robust, generalizable (accuracy, simplicity, and consistency), and explainable. We highlight explainable methods and hybrid models that are guided by the data and the law of the physical process.

1.3.2 Research Questions

This thesis addresses the following Research Questions (RQ)

(A) Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis (State of the Art)

- What are the contributions of AI-based decision support tools to I4.0?
- What are the main trends and AI models applied to PdM?
- What are the performances and limits of these technologies?
- What are the key challenges, issues identified, and future research directions in AI techniques applied to PdM4.0?
- What are the potential ethical impact rules of using AI techniques for PdM 4.0?
- What are the consequences of a loss of control of AI?

- What are the influences on human welfare and integrity? The last three questions have led to the ethical issues of AI that have proliferated in the literature over the past decades.

B) Health condition monitoring of a complex hydraulic system using Deep Neural Network (DNN) and DeepSHAP Explainable XAI.

In the literature, the most common traditional AI model used to perform predictive maintenance tasks on continuous data are DNN. These models have several advantages including their performance in classification and prediction tasks and their applications in many industries. Despite these benefits, we cannot explain the results obtained from "black box" models. Moreover, the decision-making rules used by the algorithm are of interest to decision-makers. In addition to explainability, another issue concerns the generalization of traditional approaches. Industrialists are interested in models that can be generalized and adapted to several processes. These models are required to be adapted for the purpose of capturing both local and global patterns. We present a detailed framework for Condition Monitoring (CM) based on hydraulic systems and multi-sensor data. Although autonomous systems or decision support tools using DL approaches are interesting, there are questions that require the attention of the users in their decision-making rules:

- How to predict the conditions of the components of a hydraulic system?
- What sensors need to be monitored in order to ensure the correct functioning of the system?
- What is the importance of each sensor in characterizing the state of the system?
- Is it possible to explain the results of "black-box" models or to highlight the decision rules taken by the algorithm?
- How to explain the decision-making process by the "black-box" models in a way that engenders faith in their reliability?
- How to explain the results of a Deep Learning algorithm?
- To what extent can it identify the contexts in which it is correct and fair and those in which it is not?

(C) Physical-Informed Neural Networks (PINN) and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding (FSW)

Considering that the FSW process is very computationally and time-consuming. It is difficult or impossible to simulate the whole physical time of the process or to reach the stationary regime. In addition, the framework must be able to learn the process in the transient regime only (the very beginning of the process). Furthermore, it has to be capable

of predicting the whole process duration (transient and stationary regime). Furthermore, it has to be capable of predicting the whole process duration. In this context, we are addressing a solid mechanics problem where the viscosity in the Navier-Stokes Equation (NSE) is expressed by the Norton-Hoff law which depends on the deformation rate. This is known as the modified NSE. This equation is much more complex and requires higher-order differentiation. Since the data of this process are difficult to obtain, a numerical simulation was performed taking into account several assumptions about the tool and process parameters.

- What are the approaches to simulate the Friction Stir Welding (FSW) process?
- How do determine the optimal parameters of the FSW process?
- What are the characteristics of the developed approach?
- How to introduce physical knowledge in the developed model?
- How describe the loss function?
- Why regularize or penalize the loss function through physical constraints or laws?
- The developed framework can be used to perform anomaly detection tasks?

1.3.3 Collection of Data

To address the issues of this research we have exploited 3 main data sources.

- **Data Set for bibliometric analysis:** The first data source is a textual and structured data set extracted from the Web of Science (WoS) scientific database. We retrieved 4064 scientific documents by exploiting a query that contained a set of keywords. Each document represents a record or row in the dataset. This row eventually contains information related to the author/co-author's name, the title of the article, the name of the organization, the keywords, the name of the journal, the publication date, the volume, the pages, the International Standard Serial Number (ISSN), the Digital Object Identifier (DOI), the Uniform Resource Locator (URL), and the abstract. In addition to the article information, we can also get statistics on the institutions/universities, the countries, the sources/journals, the authors, and the collaborations.
- **Condition Monitoring of hydraulic systems Dataset:** This is an experimental data set obtained via a hydraulic test bench. The sensors take the measurements of the process cyclically (once a minute). The measured values are pressures, volume flows, and temperatures. In addition, the components to be monitored are the cooler, valve, pump, and accumulator. We, therefore, have a set of 18 separate files containing multivariate data. Each file represents the data collected by a single sensor. These files contain respectively 2205 instances and 43680 attributes (numeric data). The last file represents the information (categorical data) about the conditions of degradation of the components of the system.

- **Dataset for the analysis of the thermodynamic FSW process**: The data used for the analysis is obtained through a numerical simulation performed by the FVM method. The Computational Fluid Dynamics (CFD) simulation software used is named Ansys Fluent. During this simulation, the constraints and the values of the tool parameters were fixed. In addition, we consider the entire welding process phase (dynamic and stationary phase). Furthermore, we assume that the friction coefficient is a constant, and we neglect the influence of the velocity on the friction coefficient. The resulting file is a concatenation of several files that contain 225230 rows and 7 columns (cell, 2-zone, x-coordinates, y-coordinates, x-velocity, y-velocity, and pressure).

1.3.4 Research Scope and Contributions

In this subsection, we will briefly present our contributions. To address the first objectives (see Sub-section 1.3.2 point (A)), we exploit a data mining technique known as bibliometric analysis. This method is used to study the state of the art of AI models applied to PdM4.0 and anomaly detection methods. Thus, it uses data analysis and visualization tools such as Biblioshiny, VOSviewer, and Power BI. To collect the data (4065 documents) we used a specific query applied to the Web of Science database. In addition, the query considers the publication years, title, abstract, and author/indexing keywords of the articles. This data mining technique allows us to quantify the most important concepts, application areas, scientific contributions of the methods, and thematic and main trends of AI applied to PdM4.0 and anomaly detection. The results of the analysis highlight the technological and scientific progress in the exponential exploitation of decision support tools based on machine learning models such as Deep Neural Networks. In addition, it highlights the characteristics of these approaches including their limitations and possible challenges.

Regarding the purposes of the second point (see Sub-section 1.3.2 point (B)) we exploit the main results of the bibliometric analysis. Thus, we have developed a hybrid framework consisting of two main modules. These modules are applied to the multi-sensor data fitted to the hydraulic system. The seventeen sensors collect various data such as pressure, temperature, engine power, volume, and cooling power. In this instance the problem addressed is a multi-class classification task where the data is unbalanced. To predict the different conditions of degradation of the hydraulic system components, the first module focus on a DNN model whose loss function has been regularized with a Dropout function. The resulting model is robust and efficient in predicting the component conditions (cooler, internal pump leakage, valve, condition of the hydraulic accumulator) and global conditions (the stable flag)

Despite the performance of the model DNN, we have to explain the role of each sensor (measurement variations) in the decision-making of the model DNN. For this purpose, the novelty of the second module based on an explanatory XAI method (DeepShap) coupled with the DNN results has been developed. This module highlights the mathematical decisions made by models in their training phases. Thus, the integration of explanatory approaches allows for the extraction of local or global knowledge from the DNN. For example, in this practical case, the DeepShap explainable model shows that the cooling condition of the hydraulic system is

most probably conditioned by the quantity of cooling pumped, the pressure, the power of the engine, and the temperature of the cooler to maintain the pump at a normal temperature. In addition, this approach changes the perception and confidence that we have with regard to AI models applied to industrial systems. In that respect, explainable models address several issues of ethics, transparency, partiality, and reliability of AI models, including their large-scale application.

Concerning the last point (see Sub-section 1.3.2 point (C)), we propose an optimal hybrid model based on neural networks informed by constraints or physical laws. This model is known as a Physical-Informed Neural Network (PINNs), where the problem is a regression task. The data used to train the PINN model is obtained from the results of a numerical simulation. This simulation was performed using the Finite Volume Method (FVM) with some constraints on the tools. To predict the process parameters we introduced the governing equations such as the modified Navier-Stokes Equation (NSE) directly into the neural network (NN) using automatic differentiation. The novelty is the ability of the framework to learn the FSW process at the very beginning of the transient regime and predict the whole duration of the transient and stationary regime.

In contrast to the previous method where we extracted the knowledge from the model, in this case, the physical knowledge is added to guide or inform the model. This knowledge force the model to respect certain constraints and to follow the topology of the system during the learning stages. Thus, the loss function is penalized via the physical constraints or regularization function (NSE, Dropout). By including these regularization terms, we obtain a compound loss function. which respects the properties of a classic loss function. The main results of this framework have shown that once trained, the PINN model can be a valid substitute for numerical models of thermomechanical processes to make rapid predictions. Consequently, it is also able to study various process parameters. The PINN model reduces the large computational time due to its memory effect and allows us to find an approximate solution to analytically unsolvable PDEs. The model can predict the velocity and total pressure fields and the results are in agreement with the solution of the numerical simulation. By regulating the combined loss function, we demonstrate that the resulting model is generalizable.

1.4 Organization of the Manuscript

The organization of this thesis is presented in figure 1.5. After the general introduction (see chapter 1), we will answer the research questions of this study. Thus, in chapter 2 we provide a bibliometric analysis of the application of Artificial Intelligence in Predictive Maintenance in Industry 4.0. Chapter 3 proposes a hybrid framework for predicting the operating conditions of the hydraulic system. In addition, this framework includes a model to explain or extract knowledge about the predictive model. Chapter 4 presents the Physical-Informed Neural Network (PINN) by the Navier-Stokes Equation. This model aims to predict the optimal parameters of the thermomechanical FSW process. Finally, chapter 5 is dedicated to the conclusion or review of the contributions proposed during this thesis and the future works.

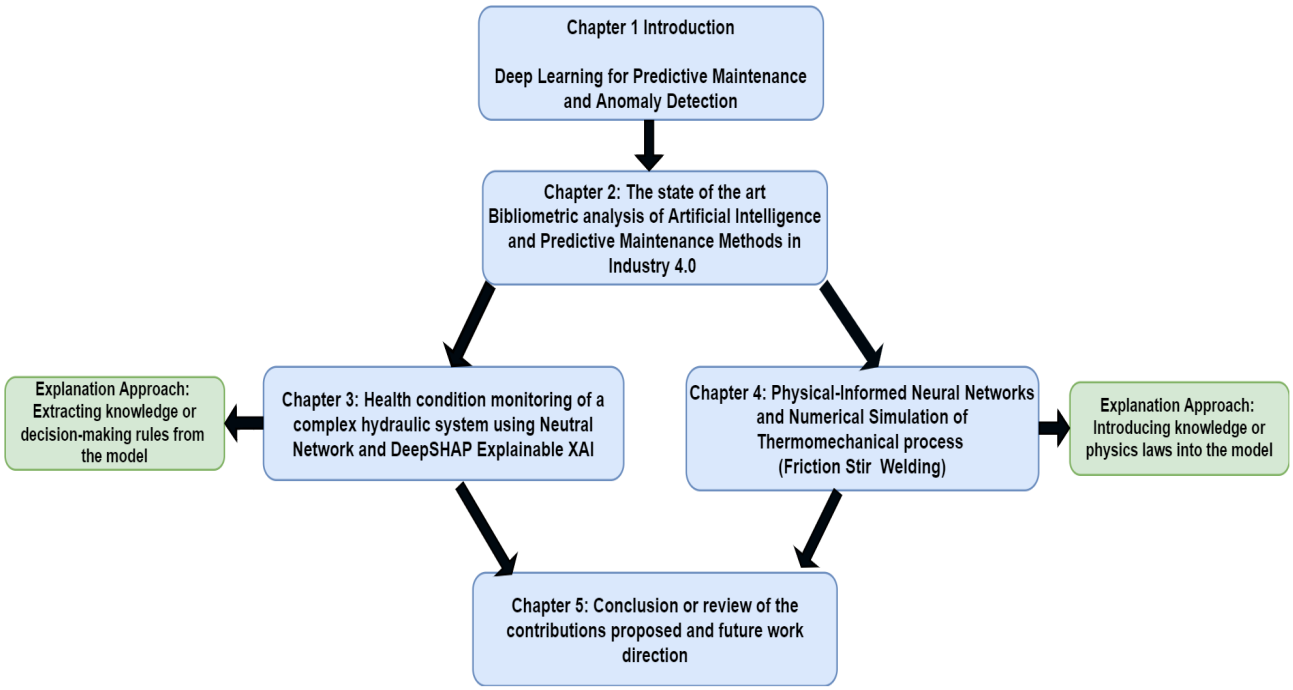


Figure 1.5: General diagram of the organization of the manuscript

1.5 Publications

This section presents the scientific articles published or under review. Furthermore, these articles constitute part of the results of this thesis.

- A. Keleko & al. Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis. AI & Ethics (2022). https://doi.org/10.1007/s43681-021-00132-6 (Published on 10 March 2022)
- A. Keleko & al. Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP Explainable XAI Advances in Engineering Software Journal (2022). https://doi.org/10.1016/j.advengsoft.2022.103339 (Published on 15 November 2022).
- Physical-Informed Deep Neural Networks and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding Submitted in Neural Computing and Applications Journal (Under review)

Chapter 2

State of the Art of Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis

Sommaire

2.1 Introduction 19
2.2 Contributions and Research Objectives of the Chapter 20
2.3 Overview of the Industrial Revolution and Maintenance Strategies . 21
2.3.1 Revolution of Industry 21
2.3.2 Industrial Maintenance Strategies 22
2.3.3 Types of Control in Industrial Maintenance 22
2.3.4 Potential Ethical Impact of the use of AI for PdM in I4.0 22
2.4 Most common AI techniques used in Predictive Maintenance 4.0 . . 25
2.4.1 Main Modeling Techniques 25
2.4.2 AI Models Applied for PdM in I4.0 26
2.4.3 Characteristics and Techniques Classifications 29
2.5 Methodology for the Study 30
2.5.1 Bibliometric Analysis 30
2.5.2 Recommended Workflow for Science Mapping 32
2.5.3 Web of Science and Data Collection 32
2.6 Analysis, and Results 33
2.6.1 Main Information About the Collection 33
2.6.2 Annual Scientific Publication Trend 35
2.6.3 Most Productive, Impact and Source Growth Dynamics 35
2.6.4 Most productive authors, universities and countries 37
2.6.5 Most common technologies or models used in Predictive Maintenance . . 44
2.6.6 Research Trends in Industrial Predictive Maintenance 46
2.6.7 Analyzes of Ethical Impact of the use of AI Techniques for PdM system . 49
2.6.8 Issues Identified, Key Challenges and Future Research Directions in PdM and I4.0 51

2.7 Discussion 52
2.8 Conclusion, Limitation and Future Works Orientations 54
2.8.1 Limitations 55
2.8.2 Future Works Orientations 55

Abstract

The purpose of this chapter consists to study the issues of industrial maintenance. Maintenance is considered one of the most important and critical drivers of Industry 4.0 (I4.0). It has contributed to the emergence of new industrial challenges. In this context, Predictive Maintenance 4.0 (PdM4.0) has seen significant progress, providing several potential advantages. These include increased productivity, through improved availability and quality. It guarantees cost reduction through automated processes and monitoring of production systems. In addition, it provides the ability to detect failures at an early stage, reduce machine downtime and predict the life of the equipment. Even though most of the exploited articles focus on AI techniques applied to PdM4.0, they do not include PdM practices and their organization. In the research work, we focused on bibliometric analysis to provide beneficial guidelines. This may help researchers and practitioners to understand the key challenges and the most insightful scientific issues that characterize a successful application of Artificial Intelligence (AI) to PdM4.0. To perform the analyses and visualize the results we used the R Biblioshiny framework, VOSviewer, and Power BI tools. These analyses highlight the most important concepts, application areas, methods, and trends of AI applied to PdM4.0. Therefore, we studied the current state of research on these new technologies, their associated methods, and related roles or impacts in developing I4.0. The result shows the most common productive sources, institutes, papers, countries, authors, and their collaborative networks. In this light, emerging topics such as Machine Learning (ML) and Deep Learning (DL) also significantly impacted PdM4.0 development. In addition, American and Chinese institutes dominate the scientific debate, while the number of publications in I4.0 and PdM4.0 is exponentially growing. This is particularly relevant in the field of data-driven, hybrid models, and digital twin frameworks applied for prognostic diagnostic or anomaly detection. Subsequently, we analyzed factors that may hinder the successful use of AI-based systems in I4.0. This includes the data collection process, the potential influence of ethics, socio-economic issues, and transparency for all stakeholders. Finally, we suggested our definition of trustful AI for I4.0.

Keywords: *Bibliometrics, Industry 4.0, Predictive maintenance, Anomaly detection, Prognostics, Condition monitoring, Artificial intelligence, Machine learning, Deep learning, Ethic, Trustful AI*

2.1 Introduction

Nowadays, manufacturers are facing increasing global competition on various strategies, and requirements such as reduction of production costs, ensuring quality and innovation of products [11]. Consequently, these manufacturers need to resort to Industry 4.0 in order to remain competitive and meet its new challenges. According to [4], the 4^{th} Industrial Revolution can be defined as a set of interconnected digital assets, and technologies that contribute to developing, automating, integrating, and exchanging real-time data in the manufacturing process. The author of [29] defines it as the integration of several technologies such as sensors, cloud computing, cybersecurity, simulation, Artificial Intelligence (AI), Internet of Things (IoT), Big data, or robotics. This new industry, therefore, meets the new requirements such as the digitalization of factories using cyber-physical systems, or communicating sensors [30]; the flexibility of the factory and the production customization [31]; the use of logistics tools that favor, and optimize the exchange of information [32]; the use of simulation techniques for configuring the production system, and making the scheduling of activities more flexible [33]. The factory must be energy, and raw material-efficient [6], and must respect some constraints such as socio-economic, ecological, and political [8]. Also, I4.0 promotes the training of the different actors [34], and the implementation of an economic strategy to be more competitive [7]. According to [1], [9], [10], the factory must be digitized to meet its new challenges. Thus, the increasing exploitation of industrial production systems, thanks to the presence of IoT, sensors, cloud computing, the widespread use of distributed control systems, and AI techniques have greatly contributed to the spread, and development of I4.0 [35]. Paper [1] shows that big data and data mining have an essential role in this development. At the same time, according to [1], [4], there are nine main pillars of technological progress that form the foundation of I4.0. Within the broad research fields related to the works mentioned above, we focus mainly on Predictive Maintenance in the context of I4.0. Predictive Maintenance 4.0 (PdM4.0) is the study of trends, behavior patterns, and correlations using some models, and real-time analysis. PdM4.0 is based on three fundamental steps (i) exploiting data collected; (ii) modeling, using different approaches among which data-driven, model-based, or a hybrid approach which combines the two previous ones; (iii) exploitation of knowledge for decision-making, and control of the physical phenomenon studied. Therefore, the resulting models allow extracting insights to anticipate breaking points and possible mechanical failures. They thus favor the decision-making process for maintenance activities in order to avoid downtime [36]. In this context, the industry can be transformed into a predictive industry [14]. Furthermore, its innovative technologies combined with machine condition monitoring systems offer new management opportunities [15], control, improvement of the efficiency, and reliability of industrial systems [16]. It should be noted that in most cases, productivity decreases are often due to anomalies or machine degradation, especially when they have not been detected. To that end, PdM4.0, machine condition monitoring, and AI has therefore become an important research area in I4.0 [37], which constitutes the focus of the present research study.

The rest of this chapter is organized as follows: section 2.2 deals with the contributions, objectives, and main issues of the study. Section 2.3 shows a brief description of the Industrial Revolution, different approaches to solving predictive maintenance challenges, and potential

ethical impacts related to the use of AI technologies, for PdM in industry 4.0. The most common predictive models used, especially AI-based modeling applied in Industry 4.0, is detailed in the section 2.4. Section 2.5 describes the research methodology used, and the process of collecting scientific publications for the analyzes conducted. A detailed and in-depth bibliometric analysis is carried out and presented in section 2.6, followed by the discussion and main contributions of the research work in section 3.7. Finally, the conclusion of the study, the limitations, and the future research envisaged are described in section 4.7.

2.2 Contributions and Research Objectives of the Chapter

Industry 4.0 and Predictive Maintenance have impacts on most aspects of the business value. In that respect, several bibliometric studies have been carried out in order to analyze these impacts. For example, several reviews concentrate on the impact of digitalization in specific sectors such as management, economics, or ecology in the literature. While [38] focuses on the different approaches and main topics related to I4.0, the author of [39] address decision-making based on system reliability in the context of I4.0. Furthermore, [40] shows the current trends of I4.0 via a comparative study with WoS and Scopus databases. The authors [41], [42] explore the elements surrounding I4.0, and their developments in the socio-economic, service industry, and management context. Also, [43] presents the challenges, and raises the relationships between sustainability, and I4.0. The authors [44], [45] focus on emerging techniques, and trends in equipment maintenance systems, while [46] presents the evolution of AI. Article [47] describes a literature review on Machine Learning for industrial applications. Authors of [48], carry out a bibliometric study mainly focused on the detection of bearing defects when using AI.

The field of industrial maintenance is vast and includes several subfamilies' maintenance methods or approaches. In our opinion, few bibliometrics studies deal with real-time predictive maintenance in the context of I4.0, which is the main focus of the present work. This targeted field of research allows us to identify potential anomalies in production to reduce machine downtime (among several other objectives). However, the development and performance of PdM4.0 systems can be hindered by several factors that we consider in our study. This chapter provides a bibliometric analysis of the different AI techniques applied to PdM for that purpose. Furthermore, the article asks questions such as: What are the current trends of the AI models, methods, or architectures used in PdM4.0? What are the impacts, the characteristics, the performances, and the possible limitations of its approaches? What are the major challenges related to the application of their method at large scales? The main contribution is to investigate which current models, methods, or techniques of AI are mostly used in the context of PdM for I4.0. We consider the following action scheme: A detailed bibliometric analysis applied to scientific papers collected on the WoS database that deals with fault detection and predictive maintenance for I4.0 was first performed. Associated analyzes and visualizations were carried out using the Bibliometrix R tool [49], VOSviewer [50], and Power BI software. We then highlighted the main trends, challenges in industrial maintenance, and the relevant methods that support conditional monitoring, fault detection, prognostic,

and diagnosis in real-time prediction. We also showed the trends in the publication of indexed documents over time. Next, we studied the current state of research on these research works considered and their roles in developing I4.0. In that regard, we identified some insightful indicators such as the most productive authors, and the leading universities (with the most cited articles), and extracted and analyzed the most frequent keywords, including the different emerging themes or technologies related to PdM4.0. We also identified the socio-economic impacts caused by the intensive use of AI-based systems applied to PdM in the industry, the issues identified key challenges and future research direction related to I4.0 for PdM4.0.

Finally, by answering the detailed above Researches Questions (RQs), we can provide a helpful guideline for researchers to better understand the research topic, the current state-of-the-art, challenges, and future directions of AI models applied to the PdM4.0.

RQ1: What are the main means of scientific publications and their frequency in the context of the study?

RQ2: What are the most productive, impact, and source growth dynamics?

RQ3: What are the most important or popular authors, journals, universities, and countries?

RQ4: What are the most common technologies or tools used in industrial maintenance, their performances, and their limits?

RQ5: What are the research trends in industry 4.0 and industrial maintenance 4.0?

RQ6: What are the potential ethical impact rules using AI techniques for predictive maintenance in I4.0?

RQ7: What are the key challenges, issues identified, and future research directions in AI techniques applied to PdM4.0?

2.3 Overview of the Industrial Revolution and Maintenance Strategies

2.3.1 Revolution of Industry

The industry has experienced four main revolutions [51]–[53]. The 1^{st} Industrial Revolution (Industry 1.0) took place between 1780, and 1860 with the creation of mechanics, the exploitation of coal, and the development of the steam engine. The 2^{nd} revolution (Industry 2.0) for the first time brought mass production at a lower cost with the introduction of electricity and the development of transport. Industry 3.0 occurred between 1970, and 2010. It highlights new information technologies, electronics, and telecommunications. Finally, the fourth revolution (Industry 4.0) was presented for the first time at the Hanover Fair in 2013

[11]. According to [2] the definition of Industry 4.0 depends on the field of application, and research. Figure 2.1 shows the main industrial revolutions and their related inspection or control techniques.

2.3.2 Industrial Maintenance Strategies

According to the European standards [54], maintenance is a combination of actions and management techniques that can be applied to ensure the correct performance of the machine over time. Figure 2.2 represents the classification strategies of maintenance; each method is described in [55]. Corrective maintenance (CM) is the action performed when a machine has faults or breaks down. Thus, there is no work until the failure is repaired. However, preventive maintenance (PM) aims to reduce the probability of failure of components. It is performed at well-defined frequencies or periods. Recently, the predictive maintenance (PdM) and Condition-Based Maintenance (CBM) strategies have attracted more attention from manufacturers [55]. Predictive maintenance is a technique to predict the future point of failure, or the lifetime of a machine component before it fails [56]. We can exploit the masses of data to train AI algorithms to optimize the production system. According to [16], [56], [57], its algorithms can detect patterns correlated with faults, failures, or detect degradation at an early stage in order to implement adequate countermeasures.

2.3.3 Types of Control in Industrial Maintenance

In industrial maintenance, there are four main types of inspection of mechanical production systems [58]. The first type which is visual inspection consists of carrying out a physical, or periodic checkup of the system (Industry 1.0). The second type is instrument inspections, which is a combination of visual inspections, and the frequent use of instruments to monitor the system's condition (Industry 2.0). Real-time condition monitoring consisting of continuous monitoring by allowing experts to give their opinions on the system status or health (Industry 3.0) is the third type of inspection. Finally, the last type is predictive maintenance which allows experts, and data scientists to exploit the data collected in order to predict the state of life of the machines.

2.3.4 Potential Ethical Impact of the use of AI for PdM in I4.0

It is widely acknowledged that AI is invading our lives. AI is creeping everywhere, from intelligent personal assistants to robotics (among the most common usages). Within the frame of PdM4.0, it can assist in cognitive tasks by providing a wide range of solutions to prevent downtime and equipment failure and even enable a system to reconfigure itself. In fact, a vital difference between the 4^{th} Industrial Revolution from its predecessor is that we are now dealing with autonomous systems, not only automation [59]. Although fascinating, autonomous

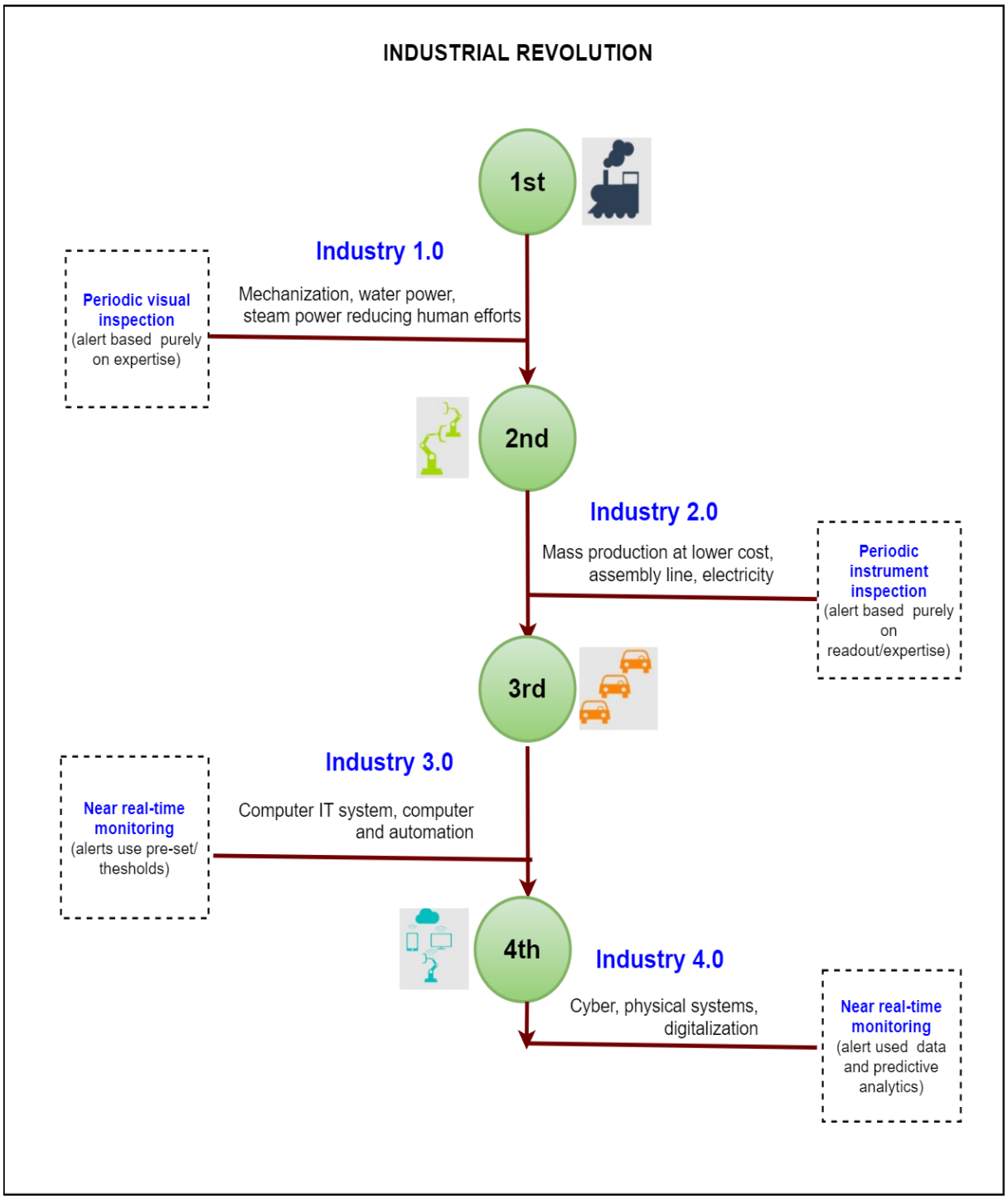


Figure 2.1: Historical perspective of Industrial Revolutions and their associated inspection or monitoring techniques.

systems are worrying: to what extent is the AI algorithm's development, outcome, and impact correct and fair? To what extent can it identify the contexts in which it is right and fair and

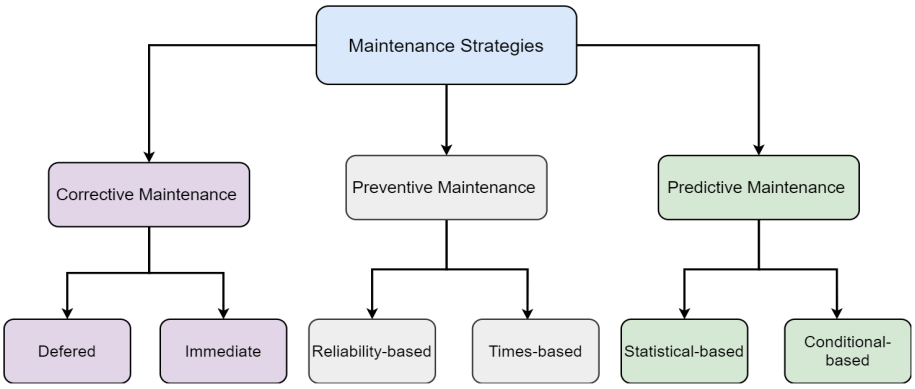


Figure 2.2: Different types of techniques or approaches for monitoring or maintenance in the context of Industry 4.0. (Adapted to [55]).

those in which it is not? What are the consequences of a loss of AI control? What are the influences on human welfare and integrity? These questions have led to the ethical issues of AI that have increased in the literature in recent decades.

As a study of what is morally wrong or right, by its very essence, following its etymology ("study of behaviors"), ethical questions define the practical principles of action. Different approaches have been developed to address the related issues, depending on the direction we give to our actions: either we act according to some moral values (virtue ethics), according to the beneficial consequences they generate (consequentialist ethics), or according to their conformity to a principle regarding some obligations, duties or rights (deontological ethics). In some cases, our actions may be subject to conflicting ethical choices, leading to ethical dilemmas [60]. The Moral Machine from the Massachusetts Institute of Technology (MIT) illustrates such a context in which an autonomous vehicle may have to choose among ethical dilemmas: saving more lives, protecting passengers, upholding the law, avoiding intervention, gender or age or species preference, social value preference as mentioned in [61]. The authors claim that exploring ethical dilemmas should be the first step to building ethical systems. Besides, while making an automated decision, [62] noted that a virtual agent could make a judgment on its ethics (individual ethical decision) or take into consideration those of other agents (within the same decision process) which may have their ethics (collective ethical decision). More recently, [63], who studied trust in AI within the field of production management, identified possible antecedent variables related to trust and which were evaluated in human-AI interaction scenarios. Their study proposed design guidelines for socially sustainable human-machine cooperation in future production management. The proposed framework is based on the SOR (Stimulus-Organism-Response) model, using decision situation characteristics as stimulus variables (predictability, error costs), AI characteristics (perceived ability, perceived comprehensibility), and human characteristics (digital affinity, expert status) as organism variables. They constructed a structural equation model in which implementation showed that AI characteristics and decision situation ones have a significant positive effect on the response, i.e., trust; for the human characteristics, they found that only one variable was statistically significant (i.e. digital affinity). Above all, following these studies and others on

ethics in AI, we believe that addressing ethical AI is a moral obligation and a duty of AI developers for PdM4.0. Therefore, we consider that ethics applied to PdM4.0 would make it possible to be proactive, support innovation positively, and not stifle its potential. Otherwise, the design of AI algorithms used in PdM4.0 may remain opaque. It can generate biases, discrimination, and worldviews without us always opening the 'black box' that makes them ethical and trustworthy.

2.4 Most common AI techniques used in Predictive Maintenance 4.0

2.4.1 Main Modeling Techniques

The main approach for anomaly detection, prognostic and diagnostic in PdM is represented in table 2.1. Knowledge-based modeling is an approach that is focused on knowledge and reasoning to solve complex problems [64]. Furthermore, this approach is based both on the conditional 'If-Then' rule, and on the knowledge known as 'Past' or 'previous' carried out in the process, also it is particularly useful to reduce the complexity of a physical model. In practice, it is often combined with other approaches as a hybrid method [65], [66]. Knowledge-based modeling can be classified into three sub-groups: rule-based [17], case-based [19], and fuzzy knowledge-based approach [18]. However, this approach is ineffective in the sense that it is impossible to apply the rules without having experience, or precise knowledge of the process being studied.
(a) Physics-based modeling requires the construction of a dynamic model by integrating various constraints, defects, or degradation linked to the non-stationary process [67], [68]. This approach has some advantages especially, since the model parameters are directly related to the physical quantities, as degradation or deformation of the phenomenon can be explained by the variations of its parameters. The results can be easily interpreted. Although the physics-based approach helps to better understand the physical universe compared to data-driven models; it is limited in its ability to extract knowledge directly from data that is mostly based on available physics. Sometimes, the models generated are often too complex leading to incorrect results [69], [70].
(b) Data-driven, or Data science modeling approach exploits both sensor data, to extract knowledge, or patterns useful for characterizing the condition of the system studied. It is based on statistical techniques, stochastic models [71], neural networks models [72], [73], data mining, and machine learning [74]. In addition, this approach is the most widely used in PdM and is a compromise between the application, and the accuracy of the model [75]. However, this method becomes unusable and loses all interest, or use when the model is no longer capable of capturing new changes associated with the process. Moreover, it does not characterize the law, or physics of the industrial process.
(c) Hybrid and digital twin modeling are a combination of a physical model, a data-driven model [76], [77], or a Knowledge-based model. Also, this approach continuously adapts to operational changes based on collected data, and online information [28], [78], [79]. Fur-

thermore, Hybrid models provide better results, especially in terms of interpretability, and understanding of physics knowledge. However, they can be costly in terms of computing time, and in some contexts, the modeling of physics can be challenging or impossible.

Table 2.1: Modeling approaches for fault detection, and diagnosis in predictive maintenance.

Modeling Approaches	**Some sub-model**
Physics-based modeling	Kalman Filters [20], Markov models [21], Monitor-based [22], Fault trees [80]
Knowledge-based modeling	Bayesian Decision [20], Expert Systems [81], Binary Trees [82]
Data-driven modeling	Genetic Algorithms [25], RF [83], Data mining [23], CNN [84]
Hybrid modeling	SAE & SVM [85], SVM & Naive Bayes [86], RF & LSTM [87]

2.4.2 AI Models Applied for PdM in I4.0

In the industrial context, AI is aimed at supporting decision-making. There are three main levels of support: descriptive, predictive, and prescriptive AI. At the first level, AI consists of providing a reliable synthesis of the massive information that is available in the form of dashboards or Key Performance Indicators (KPIs). The second level is based on a set of rules and probabilistic or statistical approaches to provide forward-looking projections in order to better predict possible risks regarding the state of the system's degradation. In addition, to providing predictive insights, prescriptive AI proposes recommendations or feedback for facilitating and optimizing maintenance operations. The models used to perform these operations can be divided into two families: Machine Learning (ML) and Deep Learning (DL). Note that depending on the nature of the explanatory data and the target variables, approaches can be classified as supervised, semi-supervised, unsupervised, and reinforcement-based learning.

2.4.2.1 Machine Learning Techniques

The Decision Tree (DT) model is an approach to represent information in the form of a tree structure with recursive partitions on the data space. DT is based on the principle of "divide and conquer", which means, the tree is built from a data set, and then it has decomposed on different subsets or branches until it reaches the last node or decision leaf (which can represent the limit of the division). Its subsets are obtained through divisions according to the Gini index. Moreover, DT is mainly composed of the main node named "root" (best predictor) among all subsets (less important predictors). The algorithm can be exploited to solve classification or regression tasks and can be used in several industry applications [88], [89]. Decision rules or results produced by the algorithm are simple and easy to understand. However, the algorithm can generate very complex trees resulting in the overfitting problem. Furthermore, DTs suffer from instability and poor performance,

compared to other ML algorithms that will be presented later.

Random Forest (RF) model has been developed by the author [90]. RF is based on combinations and aggregation (voting) of a set of random trees so that each node is evaluated independently. Furthermore, the RF model is an improvement of decision trees, particularly in the correction of the instability and the variance reduction. In addition, RF offers the possibility of extracting the significant variables involved in model construction. The parameters of the model are easy to be calibrated, robust to noise, and can be parallelized. RF is used in several applications, especially for classification or regression problems [83], [91], [92]. They are often used as a benchmark in ML competition. However, learning can be difficult (latency of the algorithm) as far as a large amount of data and a significant occurrence of missing data are concerned.

Support Vector Machine (SVM) model is developed by the author [93]. SVM deals with a generalized linear model using the hypothesis space of a linear function in a high-dimensional feature space by creating an optimal partition hyperplane (maximum distance between the bridge margins and the nearest data). Optimization problems in this constrained setting provide convex solutions. Moreover, SVM has become more popular for its applications in image classification, and face and handwriting analysis. Particularly, the authors [83], [94] apply SVM for conditional monitoring of mechanical or electronic machines. In addition, SVM uses a kernel function to guarantee better discrimination, and the regularization of the hyperparameters of the model helps to avoid overfitting problems. Some versions of hybrid SVM algorithms have been presented in [95], generally, they give higher performance than the classical SVM model. However, kernel models can be sensitive to noise data or noisy classes, to overfitting problems when selecting the optimal model. Also, the estimation of the optimal parameters can be greatly challenging since an explicit model of nonlinear kernels does not exist. Finally, the computing time or the GPU memory is important when the data to process are increasing.

K-Nearest Neighbor (K-NN) model is a non-parametric classification algorithm. Its objective is based on the classification of new sample classes with higher similarity, in this case, the K-instances nearest to the reference set are computed on a Euclidean distance metric [96]. K-NNs are very often used in industrial applications, for pattern recognition problems or recommendation systems. This approach does not require any hypothesis on the data; Furthermore, they are simple, efficient, and easy to perform. The authors [88] exploit an improved version named WKNN for fault detection and isolation tasks of complex systems. Besides, the distance-weighted k-nearest neighbors (WKNNs) are more efficient than K-NNs when the classes are separated. Nevertheless, K-NN can be inefficient because of the choice of the method of computing the distance and the number of K-nearest neighbors. Moreover, K-NNs can be inefficient due to their choice of distance computation method and the number of K-nearest neighbors. When we use a large amount of data, the algorithm becomes much slower, this is a real obstacle to applying K-NN in real-time predictive systems.

In addition to the models discussed previously, there are several classes of ML models, in particular Naïve Bayes, Discriminant Regression (LDA, QDA), penalized models (Ridge, Lasso, Elastic net), or ensemble models (Bagging, eXtreme Gradient Boosting "XGBoost").

Despite their various benefits and applications, these approaches can become unstable and inefficient (high-dimensional learning and overfitting problems) in the following cases: high data volume, complex equipment data, unbalanced classes, and missing and noisy data. Today, scientific, and technological advances have allowed deep neural network learning to emerge as a real improvement over the traditional machine learning algorithms mentioned above.

2.4.2.2 Deep Learning Techniques

Convolutional neural networks (CNN) [97] are acyclic deep learning networks, composed principally of two types of artificial neural cells: processing (convolutional) and pooling. Concerning information or feature extraction on all input samples, CNNs are based on more convolution kernels named feature extractors. To reduce the number of parameters, these kernels and weights are distributed over the entire bidirectional input matrix. CNN has shown its efficiency in various applications such as pattern recognition or signal processing. Moreover, they required very few pre-processing, since they perform their own filters during training, which explains their robustness to noisy data. However, the design of this architecture remains a major challenge for researchers. Several variants of optimized algorithms and architectures have been proposed in the literature. The AlexNet and its variant [98] is composed of five convolutional layers and three fully connected layers combined with regularization methods (data augmentation, dropout, and Norm L_1, or L_2). The Network AlexNet has won many competitions, however, it has limitations related to the image's fixed resolution, thus, the SPP network has been developed to overcome this problem. The Visual Geometry Group (VGG) network increases the depth of the network by convolutional layers with very small convolutional filters. There are other architectures such as GoogLeNet, RCNN (Regions with CNN features), and FCN (Fully Convolutional Networks). Despite their many advantages, their network (black box models) is complex, and the decision-making rules are not explainable. Besides, the increasing number of hidden layers can have an impact on the performances of the networks.

Auto-encoders (AE) are non-recurrent neural networks with hidden layers smaller than the input layers. AE is formed by an encoder and a decoder. Its objective consists of representing in an optimal way the input data. Thus, the algorithm tries to learn a new representation (encoding) from the given input data set and to reduce its dimension. To predict an output target value, the algorithm performs an optimization operation by minimizing the reconstruction error of its own inputs. Also, there are different architectures of AE. The sparse AE seeks to extract sparse features on the raw data by penalizing both hidden unit bias and hidden layer activation output. A variant named "low-density autoencoder" helps to detect objects without a priori knowledge of the class labels, the resulting model is robust to translation and rotation operations. The denoising and contractive AE have a similar network and the ability to capture details about the data. Their network structures are based on the same principle as the one shown in the previous model. Besides, denoising AE tends to introduce noise in the training set and then selects the correct information on the input of a biased model. While the "contractive" autoencoder adds explicit regularization (matrix norms such as the Frobenius norm) to its reconstruction error function, the denoising network forces the

model to learn a function that is robust to slight variations in input values. AE is efficient and has many applications, such as anomaly detection, data denoising, transfer learning, or random fake data generation. In particular, [99] use AE for the real-time remote sensing of the degradation states of the machines. Moreover, a hybrid deep SAE-SVM model is used by [85] for intelligent fault diagnosis in industry. However, the computing time of AE can be important because the problem does not prevent their exploitation of online learning.

Generative Adversarial Networks (GAN) are unsupervised learning algorithms that can generate "fake data" very similar to the original ones. The GAN algorithm is based on the game theory, where two network generators (G) and discriminator (D) are in competition. The first network is the generator of a fictitious image sample, and the second one takes the role of an adversary, checking if the data is real or from the generator. If the last one is not satisfied with the results, it returns it to the generator so that it can generate a new sample image. In addition, GANs have been the purpose of several extensions as Wasserstein GAN (WGAN) which uses the optimal transport plan to generate the data from noise, the discriminator calculates the Wasserstein distance between the distribution of the generated and real data [100]. WGAN is allowed to improve the stability of the optimization process like the search of the model hyper-parameters. Other metrics have been applied to generate or discriminate the data while improving the corresponding optimization problems, we can mention Lipschitz-GAN (LGAN), WGAN with gradient penalty (WGAN-GP), Spectral Normalization for GAN (SNGAN), First Order GAN (FOGAN), Vanilla and Least-Squares GAN. These approaches contribute to the reduction of computing time, and they are used in many applications such as pattern recognition, and generating or simulating data (texts, pictures, sounds, or videos). However, GANs are limited by the instability of unsupervised learning algorithms, and the generation of speech data is very complicated. Thus, it is not easy to turn the model training process without losing accuracy.

Finally, there are also other architectures that we have not introduced in this chapter, the Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) or the Restricted Boltzmann Machine (RBM), which are applied to sequence processing problems such as time series. However, all the approaches mentioned above are also known as "black-box" models, and their decision-making rules are not systematically explained.

2.4.3 Characteristics and Techniques Classifications

The application of AI techniques in industries can be influenced by various characteristics, including. We have the hardware and software infrastructure which provides security, the interconnectivity of systems, and information processing abilities (Edge computing and Cloud). Digital twins and decision-making help in testing the different scenarios virtually and making decisions. In this case, decision-making is based on the level of trustworthiness, and effectiveness of the model developed. Here, the evaluation and interpretation of uncertainties or error rates do not have the same significance and thus depend on the targeted objectives. In addition, an AI-based model is highly conditioned by the characteristics of the data (reliability, volume, variety, velocity, veracity, and availability). Furthermore, we could classify

these models according to some aspects: a) Nature of the task such as supervised (regression, classification), unsupervised (clustering, association), reinforcement, or semi-supervised learning. b) Type of variables to be analyzed (nominal, ordinal, discrete or continuous); c) data structure (texts, pictures, signals, videos, images, sounds); d) data quantity and quality (presence of missing, incomplete, mislabeled, noisy or biased data).

2.5 Methodology for the Study

2.5.1 Bibliometric Analysis

Bibliometric is considered the oldest bibliographic research method in information science. According to [101], it can be defined as a method for evaluating, and visualizing scientific research papers. According to [102] bibliometric analysis is a field of research that involves analyzing trends in scientific research papers on a specific topic, subject, or area. Also, bibliometric is seen as a statistical analysis applied to a set of documents, or books. Note that some organizations use this type of analysis as a distribution criterion to allocate financial aid to researchers [101]. The objective is to provide motivation, and guidance for research, or to highlight the trend, and the impact of the units. Finally, it provides motivation and guidance for research. In bibliometrics, the units of analysis frequently used are journals, documents, references, keywords, authors, and affiliations, universities, or countries and their collaborations. Keywords can be selected in relation to titles, abstracts, documents, or bodies. These keywords can be provided either by the original authors (author keywords) or indexed against referenced bibliographic data sources also known as Keywords Plus. Words represent the terms, or phrases most frequently used in the titles of the references of a scientific document [103]. Besides, they are generated by algorithms that can deeply capture the content of a document. Moreover, the authors [104] have made a comparative study between the keywords author, and the keyword Plus. Unlike the keyword author, the keyword Plus is more complex and does not necessarily appear in the title of the article. Therefore, for a bibliometric study, we can analyze several types of relations between the units, we have similarity relations, co-occurrence relations, and direct links between the units. These relationships can be represented as graphs or networks. Authors [105] present a taxonomy of the most used bibliometric techniques. Bibliometrics analysis can be applied in many fields such as logistics [106], economics, biology [107], and in industry 4.0 [40], [108]. In this article, we look for the articles using the WoS search engine according to certain criteria to avoid possible sources of error [109]. Furthermore, we evaluate the collected publications with some statistical metrics such as productivity, number of citations, frequency of citations, publications, impacts measure, and hybrid measures.

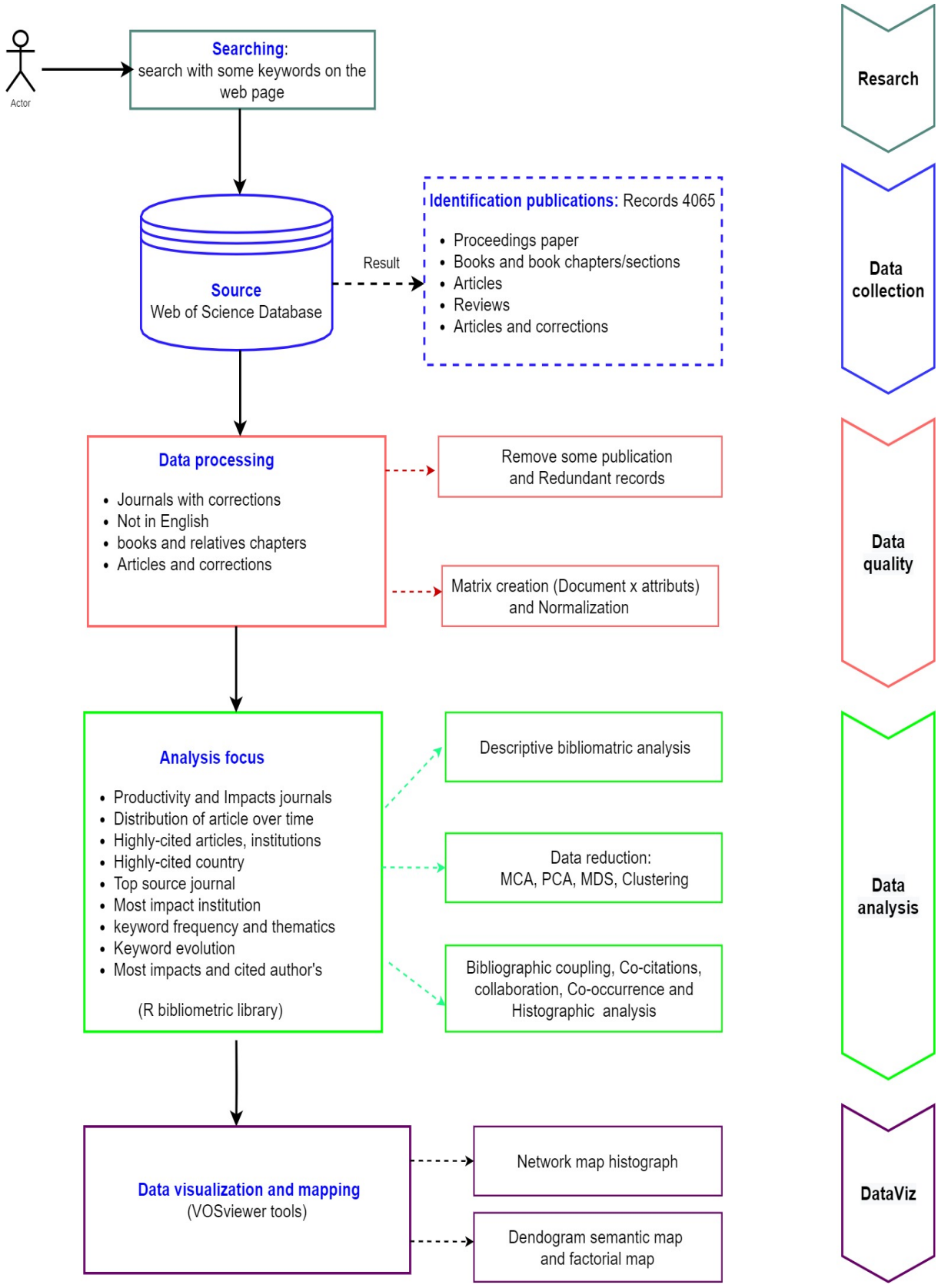


Figure 2.3: Methodology framework for bibliometric analysis. Each color corresponds to a step of methodology. The different steps represent the methods or strategies used to perform this study

2.5.2 Recommended Workflow for Science Mapping

In this subsection, we propose the four-step workflow guideline for scientific mapping research using bibliometric analysis [110]. The first step consists of defining the research questions and selecting the appropriate bibliometric methods to answer them. The second step is focused on data collection, the researchers must identify the databases in relation to the thematic study. In addition, they must perform filtering, exclusion, and selection operations to extract relevant publications. They must also consider the period to capture the evolution of the case study over time. The third step is a bibliometric analysis, which can be carried out using several statistical software [105]. The last step is data visualization, and interpretation according to the results; there are several tools available to achieve this goal [49], [105].

2.5.3 Web of Science and Data Collection

To carry out a bibliographic study we can use several bibliographic databases [111], [112] such as the web of science (WoS), Scopus, Springer, Google Scholar, or Science Direct. For our case study, we focus on the WoS search engine, our motivations are the following: (a) WoS is a bibliometric analysis tool that allows evaluating statistical indicators of publications; (b) unlike Scopus, WoS contains more multidisciplinary publications with a high impact in each field [113]; (c) in contrast to Scopus, WoS contains more multidisciplinary publications with high impact in each field, also, we exclude Scopus to avoid duplicate documents, and Google Scholar for the reduced performance compared to the quality of the search obtained. In fine, we also exclude IEEE, Science Direct, and Springer because they only index their own publications [111].

2.5.3.1 Scanning and Keywords Search

To identify important publication keywords in bibliometrics, there are several approaches [114], [115]. We applied a variant of the TF-inverse document frequency (TF-IDF) method described by [116], that helps in the identification of an important term by combining their popularity and their discrimination. This approach has several advantages, for example, TF-IDF weights are more relevant for keyword frequency than TF-KAI weights [117]. According to this index. We found that keywords such as AI, real-time, and PdM are the most important and correlated (significant increase) to productivity on I4.0. To define the relevant publication sample, we used these keywords to perform several queries on the WoS engine. The search also considers the years of publication, the title, the abstract, and the author/indexed keywords of the articles. We performed the search on 10^{th} March 2021 in the WoS database. The research produces the bibliographical data for indexed documents (4065) including some information about papers such as titles, type of article, author publications, affiliations, countries, keywords, abstracts, number of citations, source conference, publisher name, address, years of publication, volume, issue number, and a list of cited references.

(Fourth Industrial Revolution **OR** Industry 4.0 **OR** Mechanic* **OR** Real-Time) **AND** (Artificial Intelligence **OR** Machine Learning **OR** Deep Learning **OR** Artificial Neural Network) **AND** (Predictive maintenance **OR** Decision making **OR** Diagnostic **OR** Prognostic **OR** Monitoring) **AND** (Time span: 2000-2021)

2.6 Analysis, and Results

In this section, we focus on the main bibliometric analysis metrics [49]. Its metrics can be obtained on several levels such as sources, articles, authors, references, keywords, universities, or countries. We can, therefore, classify these elements by their impacts, productivity, their frequency of citations, and network collaboration. We can also visualize co-occurrence networks, the theme, and the trend of keywords. These analyzes provide new information and thus help to improve knowledge about scientific research.

2.6.1 Main Information About the Collection

Table 2.2 shows the main information about 4065 collected publications obtained on the WoS search engine according to the criteria. We have a total of 11268 keywords, and more than 450.000 authors (Author 14108, Author Appearance 18681, Author of single-authored documents 140, Author of multi-authored documents 13968, and single-authored documents 145). Also, we have 2308 source conferences, and more than 124771 references.

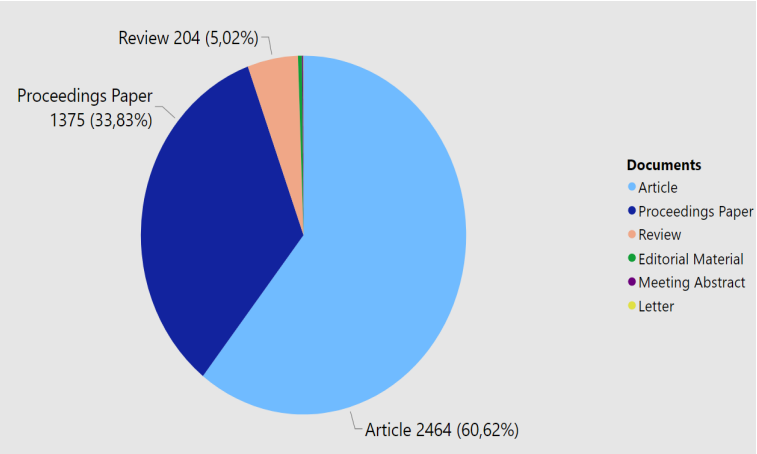


Figure 2.4: Pie Chart: The graphic represents the types of retrieved documents over the last 20 years

The pie chart (Fig. 2.4) shows the distribution of retrieved documents over the last 20 years. Firstly, we can see that articles are more representative with 2464 (60.62%) documents, secondly, and thirdly we have respectively the proceeding papers 137 (33.83%), and the reviews 204 (5%).

Table 2.2: Main information and statistics regarding the collection published between 2000, and 2021 on WoS.

Description	**Results & Statistics**
Article (2463)	- Articles (2291) - Book chapter (10) - Proceedings paper (70) - Article data paper (4)
Review (204)	- Classic review (198) - Early access (6)
Proceeding paper	1375
Editorial material	16
Meeting abstract	5
Editorial material	16
Period	Years (2000-2021)
Document contents	- Keyword Plus (5170) - Author's keyword (11268)
Author publication	- Author (14108) - Author appearances (18681) - Single-authored doc. (140) - Multi-authored doc. (13968)
Author collaboration	- Single-authored doc. (145)
Source conference	2308
References	124771
Average year of publication	3.76
Average citations per document	8.363
Average citations per year per document	1.752
Collaboration Index	3.56
Co-Authors per Documents	4.6
Documents per Author	0.288
Author per Documents	3.47

2.6.2 Annual Scientific Publication Trend

In this subsection, we answer the question **RQ1**. We defined scientific productivity as a metric that measures the frequency of publication, or author's impact on a specific discipline. Figure 2.5 shows that over the last 20 years, there is an exponential increase in the number of publications. Furthermore, table 2.3 shows the evolution reaches its peak in 2020 with more than 1095 papers published $(25,96\%)$ compared with the previous year. At the end of the first quarter of the year 2021, we record about 175 indexed documents.

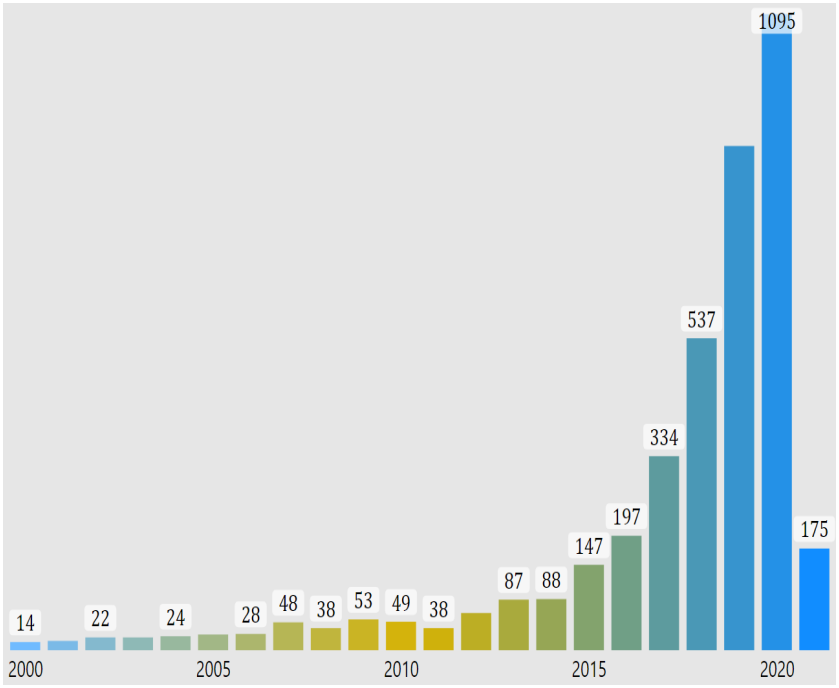


Figure 2.5: Annual scientific production published in WoS journals over the last 20 years.

2.6.3 Most Productive, Impact and Source Growth Dynamics

To answer **RQ2**, we analyze the documents, their impacts, growth, productivity, number of citations, and network collaboration. We have 2295 conferences, table 2.4 shows the most productive journals according to the Number of Publications (NP), Total of Citations (TC), and impact (h-index, or Hirsch index). We can note that H-index gives the number of publications by which the author has received at least h citations. When we focus on the NP metric, we can see that the most relevant, and productive source is the IEEE Access conference with a score of 143 (6%) papers. This journal publishes scientific papers related to electrical engineering, electronics, and computer technology. In the 2^{nd}, and 3^{rd} rank, we have respectively the Sensors applied Sciences-Basel (119), and Remote Sensing journals (47). However, the most productive source is not necessarily the most cited, and vice-versa. For example, Sensors-Basel is most cited than IEEE Access, even though, it is less productive than Mea-

Table 2.3: Productivity: Annual number of published articles between 2000-2021 on WoS. ND is number of the documents, and ND (%) is a number of documents in percent.

Year	ND	ND (%)	Year	ND	ND (%)
2000	14	0,33	2011	38	0,90
2001	16	0,38	2012	64	1,52
2002	22	0,52	2013	88	2,08
2003	22	0,52	2014	147	3,48
2004	24	0,57	2015	197	4,67
2005	27	0,64	2016	334	7,91
2006	28	0,66	2017	334	7,2
2007	48	1,14	2018	537	12,7
2008	38	0,90	2019	868	20,58
2009	53	1,26	2020	1095	25,95
2010	49	1,16	2021	175	4,14

surement journal. Regarding the source network collaboration, we consider only conferences with more than 5 publications. Finally, figure 2.6 shows the network visualization for the most productive journal, this network showed 113 conferences distributed in 16 clusters.

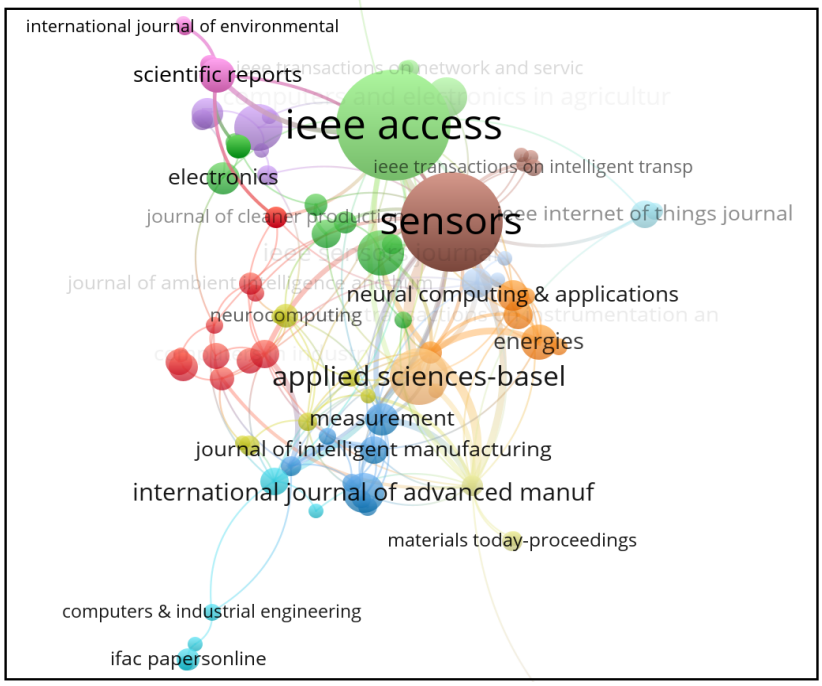


Figure 2.6: Network visualization for most productive journal.

Table 2.4: Most productive journal sorted by the publication number (NP), most impact's document (h-index), and most cited journals (TC).

Source journal	**NP**	**h-index**	**TC**
IEEE Access	143	14	814
Sensors	119	14	119
Applied Sciences-Basel	47	7	148
Remote Sensing	34	8	297
IEEE Sensors Journal	32	8	172
Computer & Electr. in Agriculture	26	8	291
Advanced manufacturing Techno.	26	6	124
Energies	21	5	142
Scientific Reports	21	6	114
Electronics	18	4	20
Measurement	18	17	303
Neural Computing and Applications	17	7	70
Plos One	17	5	166
Computer in Industry	15	7	370
Expert Systems with application	15	9	227
Intelligent manufacturing	14	7	242
Computer in Industry	15	7	370
Expert Systems with Application	15	9	227
IEEE Trans.on Instr. & Measurement	113	15	41
Multimedia tools and Application	15	4	41
IEEE Internet 0F Things Journal	14	4	132
Journal of Intelligent Manufacturing	14	7	242

2.6.4 Most productive authors, universities and countries

To answer the question **RQ3**, we exploit several axes of research, and we perform analyses to describe some elements such as authors, references, universities, countries, and continents.

2.6.4.1 Most Productive and Highly Cited Authors

Table 2.5 shows the most cited authors based on the TC index, Bellini, Filippetti, and Tassoni (694 total citations) are the most cited authors with the same score although they have published only one article. They have received remarkable attention from the community for their publication. However, if we focus on the TC index (Table 2.5), we can note the most cited authors are not necessarily the most productive. Furthermore, figure 2.7 represents the network collaboration between the authors. In this network, the distance between two authors indicates the relationship between them in terms of co-citation links. Also additionally, the link is stronger when the distance is high, or the relationship is strong. The spheres dimension is proportional to the frequency of collaboration, and the connections indicate the presence

of collaboration. We have 24 clusters, 1^{st} cluster (Gupta, Naizi and Varma), 2^{nd} cluster (Massaro, and Galiano), and 3^{rd} cluster (He, Tiwari, and Wang). Ultimately, the most co-cited authors are respectively Lecun (343), Breiman (250), He (218), Hinton (171), Lee (168), and Hochreiber (165).

Table 2.5: Most cited authors: Authors are ordered by a Total of Citations (TC) index

Authors	TC	NP	h-index	ND (%)
Bellini A.	694	1	1	0,23
Filippetti F.	694	1	1	0,23
Tassoni CA.	694	1	1	0,23
Lin J.	502	3	4	0.92
Jia F.	487	4	4	0.92
Liu C.	477	8	21	0.46
Xu X.	467	8	15	3.44
Zheng Y.	461	4	8	2.29
Lei Y.	452	2	2	5.33
Dinx SX.	419	1	1	0,23
Ozcana A.	340	4	5	1.15
Zhang Y.	390	10	18	4.20
Liu F.	383	3	4	0.92
Wang C.	381	4	13	3,74
Liu H.	378	6	19	5.01
Li Z.	363	7	18	4.20
Hsieh HP.	355	1	1	0,23
Bao Z.	349	1	1	0,23

2.6.4.2 Most Productive and Cited Affiliations

Table 2.6 shows the list of the most productive institutions. In the 1^{st} rank, we have the University of Illinois with 81 publications, in the 2^{nd}, and 3^{rd}, we have respectively the University of Shanghai Jiao Tong (74 publications), and California Los Angeles (71 publications). Table 2.7 shows the most cited organizations. Moreover, when we look at the collaborative network organization (Figure 2.8), we notice that the University of Chinese academy sciences (green cluster) is the most collaborative with 49 publications, and 684 citations, followed by Shanghai Jiao Tong University (24 publications, and 420 citations) and Georgian Institute Technology (23 publications, 410 articles). We can conclude that organizations from the USA, and China globally dominate the research in the field of study.

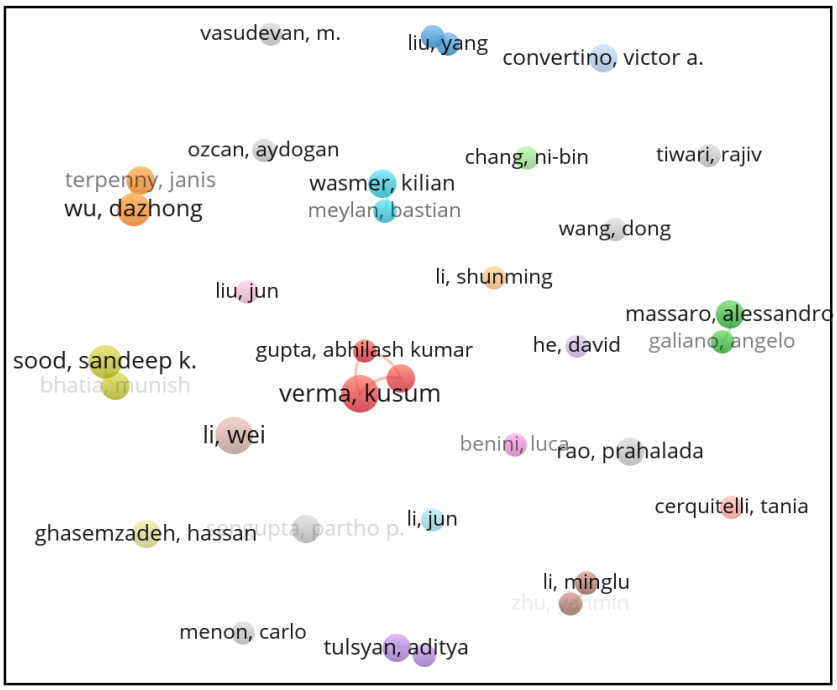


Figure 2.7: Network visualization for Publication highly Co-authorship. Each cluster is represented by a color. To interpret the results, and the color of the legends in this figure, the reader can refer to the Web version of this manuscript.

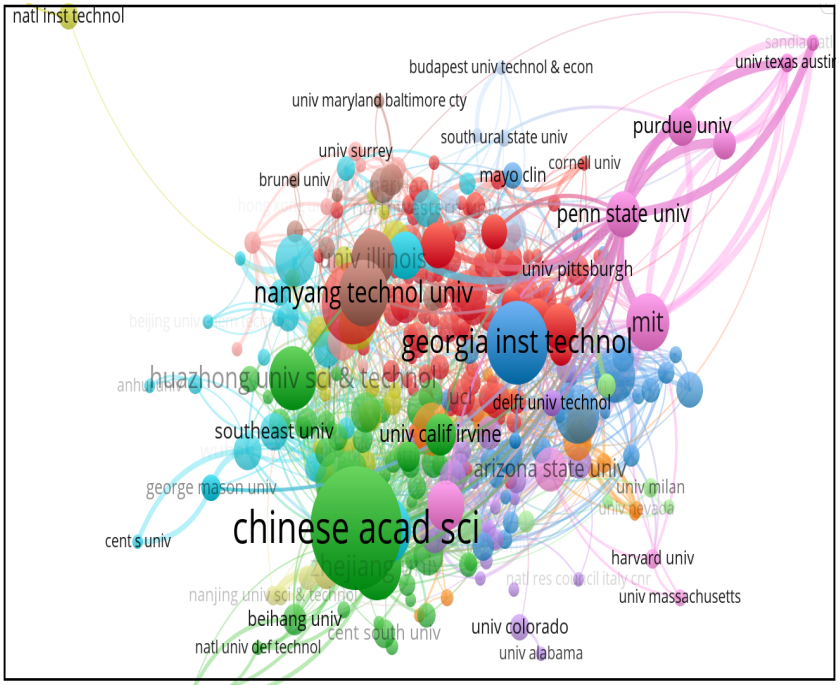


Figure 2.8: Network visualization for international collaboration affiliation. Each group is represented by a color.

Table 2.6: Most relevant affiliations ordered by a number of articles.

Affiliations	**N Articles**
University Illinois	81
Shanghai Jiao Tong University	74
University California Los Angeles	71
Nayang Technology University	66
Tsinghua University	65
Zhe Jiang University	65
Stanford University	61
Huazhong University Science and Technology	59
Xi and Jiao Tong University	51
Northwest University	47
University Michigan	46
Seoul National University	45
Yonsei University	43
King Saud University	43
University California Irvine	43
Emory University	41
Imperial College London	39
Northeastern University	39
University California San Diego	39

2.6.4.3 Scientific Productivity by Country and Continent

Regarding the country's scientific production, table 2.8 and figure 2.9 show that the USA and China are the most productive countries. We have already observed this trend when we study the most important institutes (subsection 2.6.4.2). We can, therefore, deduce that the underdeveloped countries are not representative, for example, Oceania and Africa have respectively (399) 3%, and the African continent (212) 2% publications. This trend implies that these continents are lagging even though research activities are dispersed on a global scale.

In addition, figure 2.10 illustrates the network collaboration between countries confirming that these countries are behind in research in the study field. These low productivity trends of the universities, or institutions belonging to developing, or Third World countries can be partly explained by the low collaboration between authors from developing countries. Also, the lack of infrastructure, access to digital services such as the internet, energy, and the reputation of the institution in the scientific community are factors hindering this development.

Table 2.7: The most cited organizations ordered by TC index.

Affiliations	**Total of Citations (TC)**
University California Los Angeles	822
Xi Jiao Tong University	791
University Bologna	770
University Modena and Reggio Emilia	713
University of the Chinese Academy of Sciences	684
University of the Chinese Academy of Sciences	684
Stanford University	631
Georgia Institute Technology	489
London's Global University	475
University California San Diego	442
University Michigan	439
Massachusetts Institute of Technology	420
Tsinghua University	419
Northeastern University	394
University Pittsburgh	392
University Cincinnati	347
Qatar University	339
Hong Kong University	330
University of Southern Queensland	316
Los Alamos National Laboratory	309

2.6.4.4 Most Global Cited Papers and References

Table 2.9 globally shows the most cited documents published in the WoS database over the last 20 years. In particular, paper [118] published in the IEEE Trans Ind Electron conference is the most cited (694 citations). Here, the authors are working on AI, and decision-making models are applied to the fault detection, diagnosis, and condition monitoring of electrical machines. Paper [119] has 419 citations, the authors present the application of regularized sparse filtering model for intelligent fault diagnosis under large speed fluctuations. Furthermore, the scientific article [120] is cited 355 times, and it was published at the conference on knowledge data mining.

In addition, authors Lecun [97], Breiman [90], and He [135] are the 3 most cited references (199, 169, and 156 frequency co-citations). Furthermore, Figure 2.11, shows 6 clusters of co-citations network reference. Breimman, Lecun, and He is the most representative for each cluster. In detail, the first cluster is formed by (Lecun, Hinton, and Schmidhuber), the second, third, and fourth clusters are respectively (Breiman, Bishop, Pedregosa), (Hochreiter, Kingma, and Goodfellow), and (He, Ren, Redmon). In this regard, we see that authors LeCun (644), He (596), and Krizhevsky (591) have the highest link collaboration. Finally, we can conclude that most of these articles presented in table 2.9, deal with topics related to the digitalization of industry, use of sensor data, IoT, big data, condition monitoring, anomalies

Table 2.8: Most productive and cited countries ordered by the frequency publication, or productivity (years: 2000-2021).

Region	Frequency	Average article citations
USA	3072	11,6
China	2977	8,1
India	1019	9.7
UK	639	8,7
South Korea	627	10.5
Italy	552	11,4
Germany	498	7,6
Spain	432	13,1
Canada	428	8,41
Australia	355	8,4
France	326	6,8
Japan	248	8,9
Brazil	243	6,6
Singapore	152	9,6
Malaysia	139	8,4
Switzerland	136	10,9

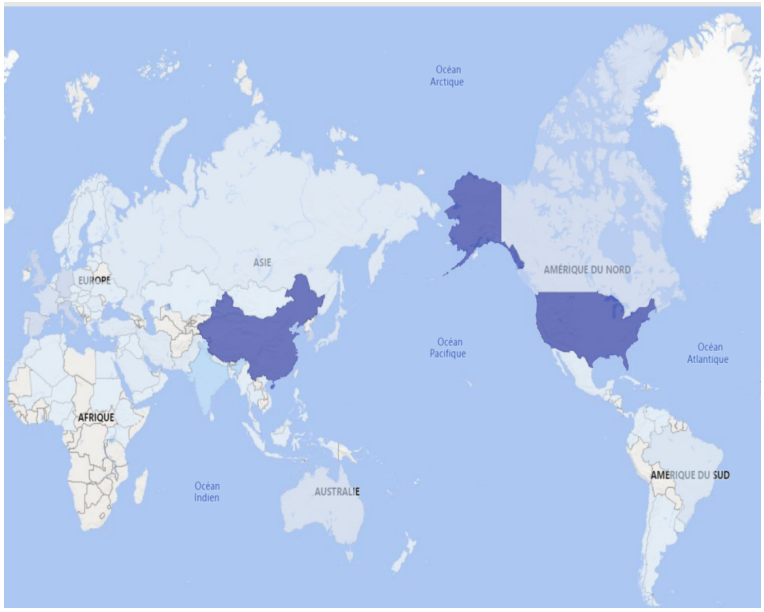


Figure 2.9: World map of the country-level scientific productivity, for documents collected on WoS over the last 20 years. The color scale is given by the number of articles, dark blue: high productivity, light blue: low productivity.

detection, ML, and DL modeling applied in PdM4.0.

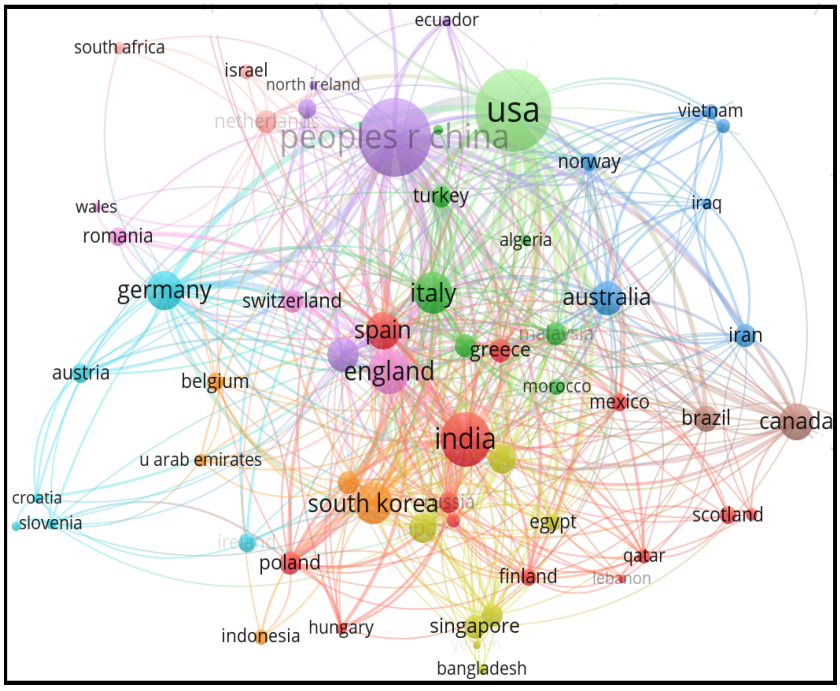


Figure 2.10: Network visualization for international collaboration.

Table 2.9: Most globally cited scientific publications.

Paper	**Frequency**	**TC/year**	**Year**
Bellini A. [118]	694	49.57	2008
Lei Y. [119]	419	69.83	2016
Zheng Y. [120]	355	39.44	2018
Benight SJ. [121]	349	38.77	2013
Mueller Kr. [122]	253	18.20	2008
Abdeljaber O. [123]	249	49.80	2017
Bigio IJ. [124]	237	10.77	2000
Verrelst J. [125]	231	15.20	2012
Khan S. [126]	225	56.25	2018
Yaseen ZM. [127]	203	67.66	2010
Berg B. [128]	190	28.42	2015
Oresko JJ. [129]	197	16.41	2015
Jing L. [130]	196	39.20	2010
Gonzaga JCB. [131]	191	14.69	2009
He J. [132]	191	38.00	2017
Michie S. [133]	190	38.20	2017
Botu V. [134]	170	24.28	2015

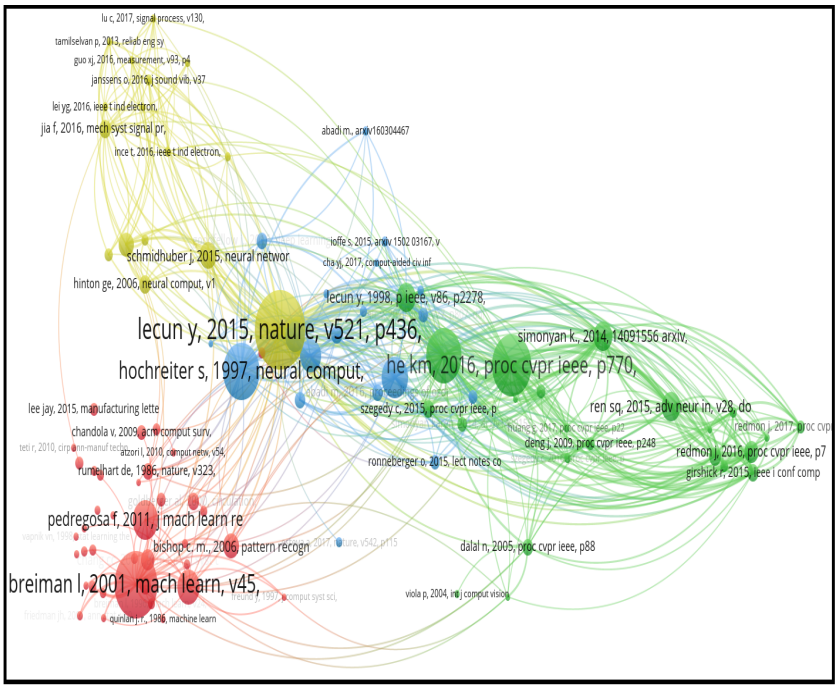


Figure 2.11: Network visualization for Co-citations. Cluster 1: yellow color, cluster 2: red color, Cluster 3: green color, and Cluster 4: blue color.

2.6.5 Most common technologies or models used in Predictive Maintenance

2.6.5.1 Most Frequent Keywords and Co-occurrence Analysis

In this section, we focus on the question **RQ4**. We first analyze the most frequent keywords and their co-occurrence networks. The co-occurrence network keyword is a relational bibliometric metric frequency of scientific knowledge. So, the node represents a keyword, and its size is proportional to the frequency of co-occurrence of the word. While the color determines the cluster to which the element belongs. Thus, its clusters provide a global view of divergent research areas and group words according to the scientific field of research. Moreover, two keywords tend to be relatively close when they appear more frequently in the same articles. Furthermore, the distance between two nodes in the figure is determined by the density. To improve the analysis, we considered the most frequent keywords in each group and the keywords that appear at least three times in the abstract. Table 2.10 presents the list of the most frequent author keywords in the publications. We can see that the most frequent author's words are machine learning with 792 occurrences followed by deep learning, artificial intelligence, and monitoring with respectively 479, 286, 220, and 177 occurrences. Finally, the 2.12 shows the density of the author keywords co-occurrence, and table 2.11 represents their word clustering. This illustrates the most important keywords, machine learning, fault, diagnostic, intelligent systems, data science, CNN, ANN, computer vision, network monitoring, or on-line

monitoring, have a great impact or importance for I4.0 and Pd4.0. In particular, DL and ML approaches have a major role in solving PdM problems in I4.0.

Table 2.10: Top 20 of the most frequent authors' keywords.

Author's Keyword	**N occurrences**
Machine learning	792
Deep learning	479
Artificial intelligence	286
Monitoring	220
Artificial neutral networks	177
Learning	166
Machine	108
Internet of things	792
Classification	479
Fault diagnosis	88
Feature extraction	85
Sensors	83
Big data	82
CNN	78
Industry 4.0	75
IoT	71
Condition monitoring	66
Predictive maintenance	64
Anomaly detection	62
Real-time	55

Table 2.11: Cluster of co-occurrence network author's keyword

N cluster	**Node**
Cluster 1	Deep learning, CNN, structural health, neural networks, computer vision ANN, object detectors, LSTM, real-times, monitoring, transfer learning
Cluster 2	Monitoring, fault diagnosis, sensors condition monitoring, real-time systems, Signal processing, forecasting training
Cluster 3	Machine learning, IA, IoT, Big Data, data mining, pattern recognition classification, I4.0, anomalies detection, RF, predictive maintenance Health monitoring, edge computing,
Cluster 4	Support vector machine, remote sensing, neural networks, image processing

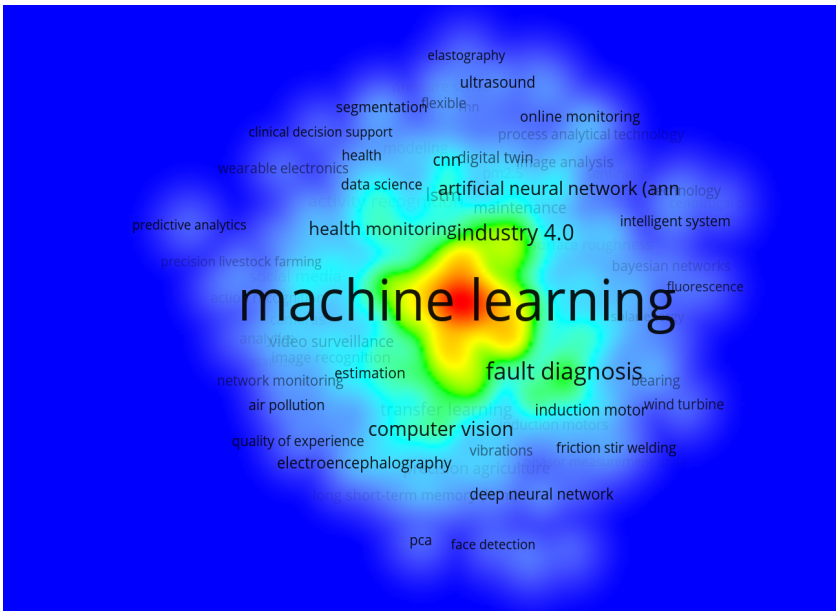


Figure 2.12: Density visualization map of the most frequently related terms in retrieved articles on WoS. The frequent terms were visualized using VOSviewer.

2.6.5.2 Keywords Conceptual Structure Map

Secondly, we use Multiple Correspondence Analysis MCA to analyze the keyword conceptual structure Map. MCA approach is an exploratory multivariate technique for the analysis of multivariate categorical data [49], [136]. We can explain the importance of the keywords in relation to their positions on the map, and on the main axes. Also, the proximity between two keywords implies that they have a similar distribution. Figure 2.13 shows the distribution of the most common keywords with the minimum number of documents (10) grouped into two groups. We can notice that keyword such as big data, sensors, diagnosis, systems, or simulation is located on the same plane and are very close. Furthermore, keywords such as IoT, and the internet belong to the second axis. These keywords are, therefore, associated with the most common technologies applied to the PdM in I4.0. Lastly, the performance and limitations of these models have been described in section 2.4.2.

2.6.6 Research Trends in Industrial Predictive Maintenance

2.6.6.1 Keywords Dynamics Analysis and Trend Topics

Regarding the question **RQ5** we analyze several elements. Initially, we investigate the keyword evolution associated with the topic study (Figure 2.14). From 2014, we note a real emergence of the use of approaches such as AI model-based (DL, ML) applied to monitoring, diagnostic technique, and PdM4.0. When we focus on keyword plus, we have terms like clas-

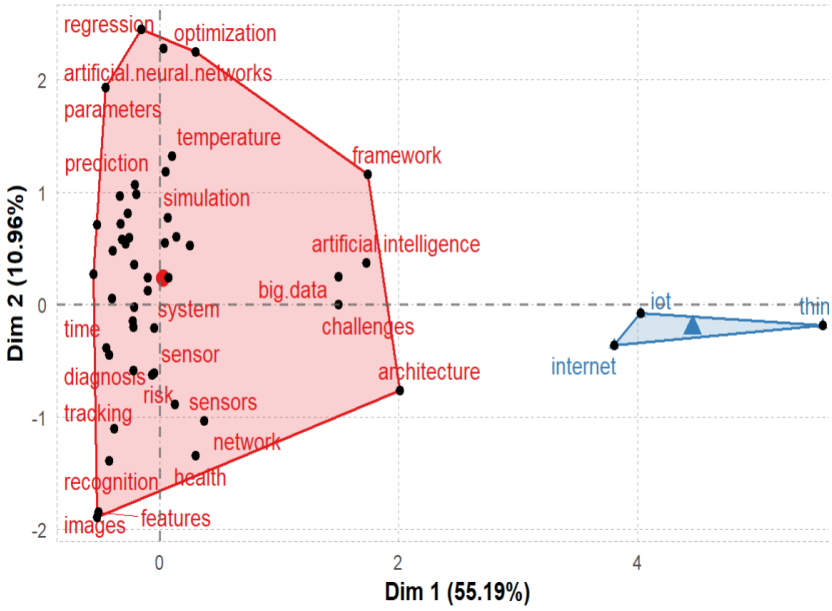


Figure 2.13: Conceptual Map, and keywords clusters (minimum number of documents is 5, and method is MCA. Each cluster is represented by a distinct color (Cluster 1: red color, and cluster 2: blue color).

sification, systems, data, real-time analysis, prediction, identification, and diagnostics that are important to develop anomaly detection, condition monitoring, and PdM4.0 systems.

Furthermore, figure 2.15 shows the topic trend of the collected documents over the last 20 years. From 2018 we observe an increasing use of several technologies and models approaches such as sensors, fault detection, condition, health monitoring, data, IoT, and data-Based Modeling that support the rise of PdM4.0 and I4.0. Also, we have observed this evolution and development in figure 2.14.

2.6.6.2 Thematic Map and Thematic Evolution

An additional element that guides us to answer the question **RQ5** is to use a co-word for analyzing the evolution or trend of the most significant research thematic. Regarding the co-word analysis, each cluster represents the different conceptual themes developed in the domain and the research period. Thus, authors [137] defined a strategic diagram by Callon's centrality metric, which measures the degree of relationship, or links between each cluster. In addition, the strength, and the number of links imply a major relationship between the research problems in the scientific community. Indeed, Callon's density measures the strength of the links between the keywords of each cluster or evaluates their impact over time in the network. Lastly, the volume of the spheres is proportional to the frequency of publications associated with each research theme. Figure 2.16 shows a strategy graph that represents the search sub-clusters in a bidirectional space.

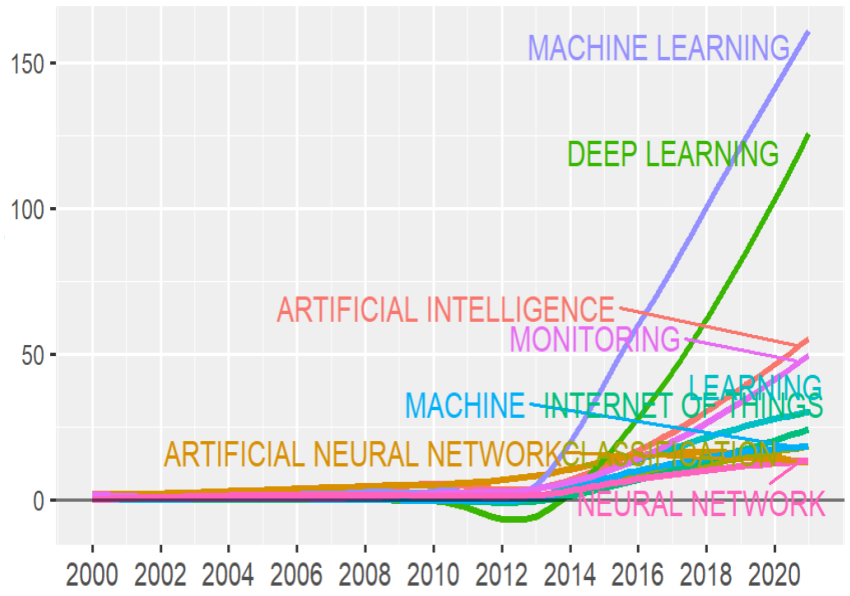


Figure 2.14: Word dynamics for the keyword plus

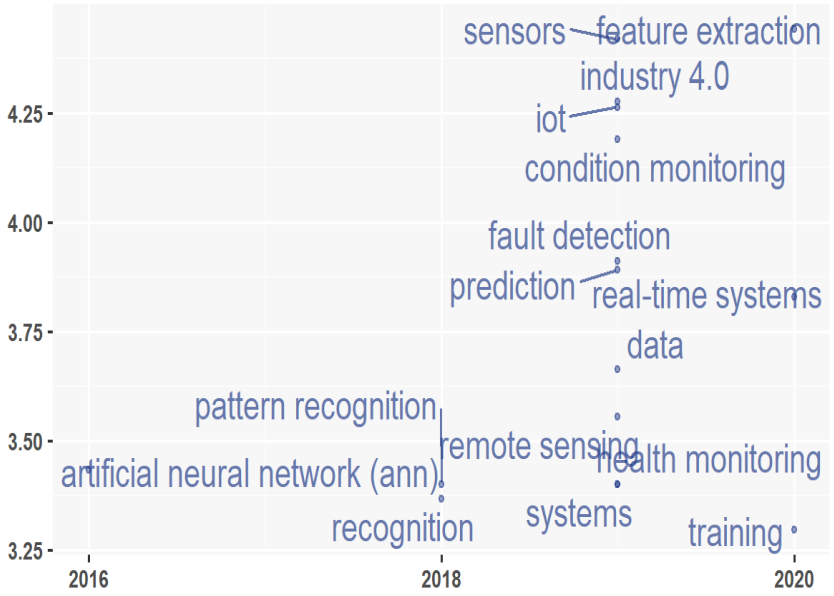


Figure 2.15: Topic trend analysis of the collected documents over the last 20 years.

Regarding the results, we consider the 200 most frequent keywords described in relation to author keywords. Figure 2.17 represents the strategic maps of the main thematic and trends topic. According to this figure, we have six main thematic (Industry 4.0, artificial neural networks, monitoring, deep learning, and machine learning) for the author keywords.

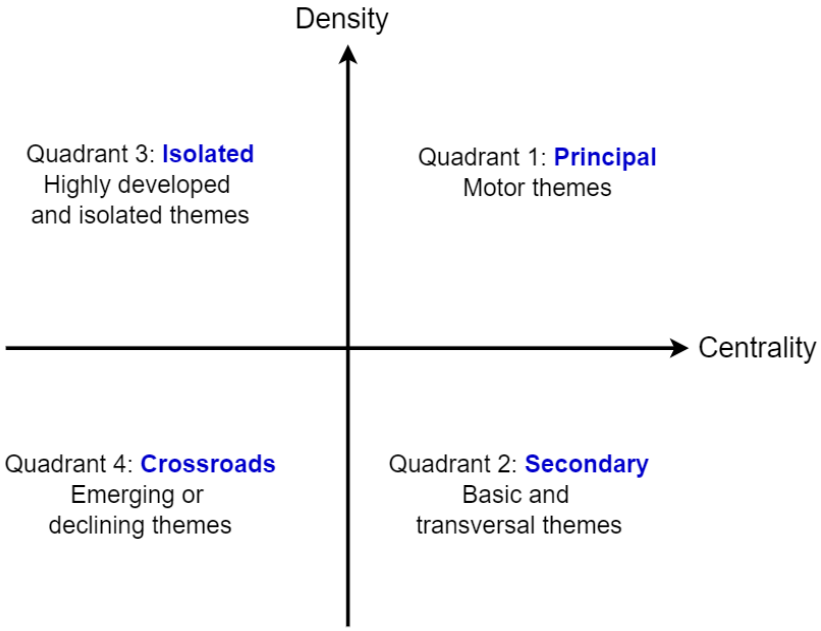


Figure 2.16: Strategic diagram (adapted from [137])

However, when we focus on keywords Plus, we mainly have four themes (the internet, neutral networks, system, classification, and networks). Moreover, the keyword abstract or keyword titles extracted from the article's contents give three major thematics (monitoring, real-time, and data), and (times, learning, and monitoring). Furthermore, table 2.12 shows the main emerging and motor topics related to the study case as well as their corresponding subgroups topics. We can deduce that I4.0 is an emerging or crossroad topic while monitoring technique, ML, and DL approaches are the principal or motor topics. Finally, we can deduce that in recent years scientific research has focused mainly on these cited subjects or topics.

2.6.7 Analyzes of Ethical Impact of the use of AI Techniques for PdM system

Ethical issues were initially not mentioned in the constitution of the initial query presented to the subsection 2.6.7. By performing a sub-query with the following keywords: "Ethical", "Artificial Intelligence", and "Industry 4.0" on the set of 4065 papers initially collected, we found a subset of 37 papers that address the ethics and trust based on AI models. Regarding the answering to the question **RQ6**, we exploited the results of this sub-query and the probable impact presented in the subsection 3.3.2. We can therefore show that AI systems and the industry's robotization will probably have several impacts on ethics, confidentiality, privacy rules, transparency, and human-robot collaboration. Furthermore, industrialization may involve social-economic issues, particularly the increase in the unemployment rate, reduction of the workforce, and the evolution of the disparity between developed and Third World countries.

Table 2.12: Strategic Map author's keywords. Each cluster is represented by the main theme and its positioning in relation to the current literature.

Author's Keywords	**Position**
Cluster 1: Industry 4.0 RF, IA, Monitoring, PdM, System, Cyber-physical object detectors, real-time, Condition monitoring	Crossroads (Emerging theme)
Cluster 2: Artificial Intelligent Neural networks, Recognition, Prediction, Diagnostics, Image & signal processing, Genetic algorithm	Crossroads (Motor theme)
Cluster 3: Monitoring Sensors, Real-time, System, Forecasting, Neutral networks	Principal (Motor theme)
Cluster 4: Deep Learning LSTM, Feature extraction, CNN, ANN, Fault, Health monitoring, Computer vision	Principal (Motor theme)
Cluster 5: Machine Learning IoT, Big Data, Data mining, Remote, Anomalies detection, SVM, RF	Principal (Motor theme)

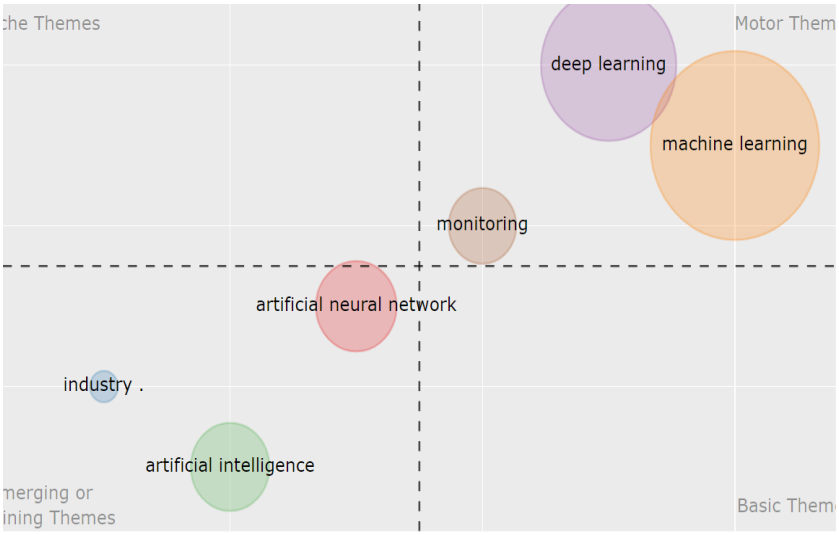


Figure 2.17: Strategic map of the author's keyword. To interpret the results, and the color of the legends keyword in this figure, the reader is referred to the color.

2.6.8 Issues Identified, Key Challenges and Future Research Directions in PdM and I4.0

To answer the question **RQ7**, we showed the challenges associated with the deployment of AI Systems in I4.0 can be associated with several factors such as operational, organizational, technical, data collection or processing, cybersecurity, interpretability, trust, privacy, and ethical rules.

(a) Operational and organizational: The growth and industrialization of the factory generally lead to reforms and changes in the human, operational, management and organizational levels [138]. These operators must be able to interact with professionals from other fields (multidisciplinary). Besides, companies must surround themselves with specialists (data scientists or experts) in the fields in which the solutions will be applied.

(b) Machine-to-machine and human-to-machine interactions: It is essential to ensure that AI systems do not affect the functioning of other equipment or interconnected machines in the production process. Thus, the industry should ensure that AI systems can interact or communicate with other devices while maintaining their behavior. Furthermore, workers must be trained or adapted to communicate and interact with these new technologies.

(c) Cybersecurity and privacy: The exponential exploitation of interconnected technologies or storage systems such as IoT, sensors, databases, or big data infrastructure (local or cloud) can expose AI systems to cyberattacks notably through spamming or malicious software classification [139]. However, considerable efforts are being made to enforce ecosystem standards and guidelines such as the ISO/IEC 29180:2012 standards for sensor networks. Nowadays, there is really no standard or reliable process to ensure the security of AI models against attacks. Further questions are raised about the General Data Protection Regulation (GDPR).

(d) Real-time data collection: Data is a key element in PdM, these data must be massive, secure, available, accessible, and qualitative to perform a generalizable PdM system. Data collection is therefore a major challenge for companies since the sensors or machines do not often generate representative data on their conditions, deterioration, or configurations. In so doing, a possible solution is to label all the raw data although this operation can be time-consuming, fastidious, and requires the knowledge of an expert. Furthermore, labeling operation involves risks of errors and considerable economic and operational costs. Indeed, to address data quality issues, several approaches have been proposed, such as artificial resampling, interpolations techniques, semi-supervised learning, or data augmentation. Thus, when data are scarce, it is also possible to use the GAN model to simulate reliable data [100], or transfer learning (different working conditions and machines) for the transfer of knowledge from one system to another [140]. Approaches cited are not systematically efficient, the data simulation process is sometimes not adapted to the real conditions of a machine's operation since the imposed scenarios do not represent the complexity of the system (machine degradation or failure).

(e) Adaptability of prescriptive and hybrid modeling in real time: it is important to

develop prescriptive and hybrid models as a recommendation system for diagnosis, prognosis, and anomaly detection of machines in real-time. Furthermore, hybrid models [141] have the benefit of integrating both physical and numerical knowledge or constraints of the system combined with data-driven modeling. Furthermore, an AI system must be able to adapt to the whole system while maintaining its performance. Adding new equipment (machines) should not be an obstacle or a constraint that may impact the model's quality.

(f) Using computer vision, multimodal prediction, images, video processing, texts, and sound data in the PdM: Data are collected from heterogeneous systems and are of diverse nature. The challenges of PdM are to combine all these data to perform a multimodal prediction.

(g) Explainability XAI-model: We have shown that some AI black box models such as CNN, or RF have a real interest in PdM problems. However, they are not easy to interpret and are neither intuitive for all stakeholders. In this context explainability is an important factor for the acceptance of AI solutions. Furthermore, a new trend of AI is focused on XAI explainability model-agnostic methods [142] such as Shapley Additive exPlanations (SHAP) [143] or Local Interpretable Model-Agnostic Explanations (LIME) [144] that are designed to explain and understand the "black-box" model decision-making, and make it easily interpretable, comprehensible and user-friendly for all stakeholders. Indeed, AI systems should not substitute humans, but support them in taking over low-level thinking tasks. In this regard, experts must collaborate with these new technologies to ensure productivity.

2.7 Discussion

This article focuses attention on a detailed bibliometric analysis based on using AI techniques for PdM in I4.0. The main objective is to highlight the evolution, impact, and current state-of-the-art scientific research related to the exploitation of these technologies for anomaly detection, diagnosis, and PdM4.0. The results obtained give us a detailed analysis to address all the questions initially formulated. Furthermore, these results show a relative description of several metrics, including the publication trend, most productive journals, documents, authors, co-authors, references, affiliations, countries as well as network collaboration between authors or institutions. We have represented the most important keywords, conceptual, intellectual, and social structure of the research, including all past, principal, and emerging themes. Furthermore, we present the potential ethical impact rules using this AI system. Besides, we discuss the main challenges, and future research directions in AI applied to PdM4.0.

We analyzed the main information about the 4065 collected documents according to their dynamics, productivity, total citation, impact on the community, and network collaborations. We have observed an exponential increase in the number of papers published in the last 20 years. We have 2308 source journals, however, the most productive are respectively IEEE Access (143), Sensors (119), and Applied Sciences-Basel (47). Indeed, we have shown that these journals are not necessarily the most cited. To have a high profile, and reputation in the scientific community, its journals are also open access, making it easy to view articles online.

Concerning most cited, co-cited authors, as well as their collaboration network, and their impacts (subsection 2.6.4.1), we noted that Bellini, Filippetti, and Tassoni (694) are the most cited authors with the same TC metric. Furthermore, we analyzed affiliations, based on their productivity, impact, and collaboration. The most productive universities are respectively Illinois (81), Shanghai Jiao Tong (74), and California Los Angeles (71) university. As far as an international collaboration between authors, or institutions, is concerned, institutions belonging to developing countries are not representative, notably Oceania with 399 articles (3%), and the African continent with 212 (2%) of the published articles. This trend of low productivity of universities, or institutions in Third World countries can be partly explained by the fact that they are lagging in scientific research due to the lack of infrastructure, access to digital services such as the internet, energy, and the reputation of institutions in the scientific community. We can conclude that the institution in the USA, and China globally dominate the research on the topic.

Regarding the most cited, and productive articles, we have the following papers: [87], [118], [119], [145]. In particular, article [118] has been cited 694 times, the authors exploit AI models for monitoring and detection of electrical and mechanical defects. Article [119] which has been cited 419 times, exploits sensor data, and DL technique for the intelligent diagnosis of failures via regularized neural networks. Author [145] uses wavelet analysis and ANN models to predict the weld quality in friction stir welding. Paper [87], presents a hybrid model (RF combined with LSTM) for real-time monitoring, and corrective adjustment. In fact, the topic described in these papers deals with subjects related to the digitization of industry by using the IA system. Moreover, the most cited references include [97], [135], [146], [147].

To identify the most discussed topics and common technologies used in PdM, we extracted more than 11268 authors' keywords. Thus, we have analyzed these words including their co-occurrence networks. We can deduce that words such as ML, DL, AI, monitoring, artificial neural networks, data science, I4.0, IoT, sensors, big data, fault diagnosis, feature extraction, CNN, condition monitoring, predictive maintenance, anomaly detection, real-time are the most frequent keywords. Moreover, the evolution of its keywords is associated with the main theme of the study. We mainly have 6 topics; however, the principal and emerging themes are respectively monitoring, I4.0, DL, and ML techniques. Besides, we have a real emergence of these approaches applied to monitoring, diagnostic technique, and PdM4.0. Indeed, the analyzes suggest that the heterogeneity and the link between these keywords reflect the importance of AI techniques in PdM4.0.

Additionally, we have observed that industrialization and automated systems probably have an impact on the whole ecosystem of the industry. AI models applied to PdM systems have some benefits, maintenance cost, and energy consumption reduction. Furthermore, they help to improve quality, to optimize, and increase the efficiency or flexibility of production processes. On the human side, these models' impacts can be organizational, operational, security, trust, socio-economic, or legal. Indeed, authors [148] show that AI systems will increase the workload of employees and create a need for adaptation and dependence on new technologies, particularly the challenge concerning transparency, and human-robot collaboration

raises ethical issues. In fact, [149] demonstrated that the transition to robotization of manufacturing systems is going to involve social-economic problems (generalized social exclusion) and a reduction of the human workforce (massive unemployment). Also, the operators will have to re-adapt, be formed, and specialize in these new challenges or operational changes. Therefore, if the public authority does not adopt actions, industrial automation and AI system use will contribute to increasing the gap disparity between technologically advanced countries and under-developed countries.

2.8 Conclusion, Limitation and Future Works Orientations

The contribution of this chapter is to provide a useful state-of-the-art basis for the literature search on the use of AI techniques applied to PdM in Industry 4.0. To address the research problem (**RQ1 - RQ7**), we have performed a bibliometrics analysis using Biblioshiny, VOSviewer, and Power BI tools. This detailed analysis is based on 4096 scientific documents collected between 2000, and 2021 from the WoS Database. We focus on some metrics, including the publication trend, most productive sources, papers, authors, co-authors, references, affiliations, and countries as well as network collaboration between authors or institutions. Furthermore, we analyze the most important keywords and the principal or motor theme associated with this study. We also analyze the benefits of AI models in the industry, their particularities, applications, impacts, and major results or performances. Particularly, we were also interested in ethical, trust, transparency, and socio-economic impacts that could be caused when using these models. We give our definition of trustful AI for I4.0 and its effects. Finally, the potential limitations, key challenges, and future research directions of AI systems are presented.

The results obtained showed a progressive increase in the frequency of publications over the last 20 years. Regarding the sources, we have shown that the IEEE (Institute of Electrical and Electronics Engineers) access is the most productive and cited journal. Moreover, the most productive universities are respectively Illinois, and Shanghai Jiao Tong University. The USA and China are the countries with a major impact on scientific research related to the study topics. Indeed, the collaboration between developed and Third World countries is very weak, while the international collaboration among developed countries is strengthened. For the author's analysis, we have observed that the most cited author and reference are respectively Bellini and Lei. Furthermore, the analysis of the collaboration networks shows that some authors tend to work in small groups (three collaborators by group), which implies a large number of groups or clusters of authors. According to the author's keyword analysis, we show that the most important theoretical knowledge and research theme on PdM4.0 are mainly in the areas of machine learning, and deep learning, including their sub-models. Moreover, we have 6 main topics among which the emerging themes are DL, ML, and monitoring. These different results clearly show a wide field of applications (monitoring, diagnosis, prognostic, anomalies detection) or different situations, especially for supervised, unsupervised, or semi-supervised learning problems.

We have described the most common predictive models used in PdM 4.0. Despite their performance and application in many industrial cases, in practice, we have shown that some predictive models have several limitations, especially on their instability and overfitting in a situation for missing or noisy data, high volume, complex and unbalanced classes. In addition, they can have complex architectures, resulting in a significant requirement for GPU, and computing time, in the estimation of these parameters. real-time or online analyzes can become very complicated due to the high computing time and complexity of these models. Moreover, these models can interfere with the correct functioning of the system. Besides, most of these "black-box" models are not explainable, i.e., the algorithm decision-making process is unknown. This can be a real issue for their generalization in the industry.

Finally, using AI technologies in the industry can also be confronted with some challenges such as operational, organizational, adaptability, machine-to-machine, human-to-machine interactions, cybersecurity (risk attacks), analysis online, real-time data collection, and data quality. Also, we have challenges concerning prescriptive, hybrid, and multimodal modeling, visual reasoning, socio-economic, Explainability XAI, interpretability, trust, privacy, GDPR data protection statements, and ethical impacts.

2.8.1 Limitations

Concerning the main limitations, we performed a search with selected keywords according to the study context. However, we cannot guarantee that these keywords and the scientific documents collected cover the whole research area. Moreover, we use an open-access journal WoS database which does not contain all the publications. Also, the scope of this research is limited to English documents collected from WoS and we used a traditional bibliometric approach to perform analyzes, therefore, by combining the different methods we can considerably improve the results.

2.8.2 Future Works Orientations

In order to improve the results, we can refine the query by including more accurate keywords. The exploitation of several bibliographic databases such as Scopus, Springer, Google scholar, Science Direct, and IEEE, as well as the selection of documents supplementary by including books, notes, and thesis, will also be applied to retrieve all documents covering the field of study and improve the quality of analysis. We will also consider contributions written in languages such as French, Chinese, Italian, Spanish, or German. Another area of improvement is to use a combination of several bibliometric analysis methods to strengthen the result.

In the following two chapters, we exploit the traditional AI models extracted from this literature review. The most commonly exploited models for solving predictive maintenance tasks with continuous data are deep neural networks. These networks have the ability to extract information via their hidden layers. Moreover, they are generally performing well in solving classification and prediction tasks. Despite their considerable success in industrial

applications, we can not fully explain the obtained results. This disadvantage is also important for people interested in understanding the choices of algorithms to discard one feature over another (decision rules). Another problem concerns the generalization of these models. Manufacturers are interested in efficient models that can be adapted to several processes. Moreover, these models must have the ability to capture both local and global phenomena. To address the issues of traditional AI approaches such as DNN, we develop hybrid models, which are usually composed of several models.

To address the first issue related to "black box" models we will integrate further knowledge to explain the results of the prediction model. This approach consists in introducing an explainable XAI framework combined with the prediction model. The framework explains the decision rules when training the developed model. This knowledge has the advantage of modifying our perception of the results of intelligence models. In addition, the XAI method can be used to respond to feature selection operations. This makes the predictive model more robust and accurate. We will give more details about this hybrid model in chapter 3.

Regarding the second challenge, we will develop a hybrid and generalizable model via the integration of the topology knowledge of the process. In contrast to the previous case where we extract the knowledge, in this case of the PINN model we add the knowledge to train the model. This forces the model to follow conditions well-defined by the laws or constraints of physics. This physics constraint modifies the loss function. The optimization of this new combined loss function promotes or guarantees robust and accurate results. This model and these key values will be presented in detail in chapter 4.

Chapter 3

Health Condition Monitoring of a Complex Hydraulic System using Deep Neural Network and DeepSHAP Explainable XAI

Sommaire

3.1 Introduction 58
3.2 Literature Review of Condition Monitoring Applied to Hydraulic Systems 61
3.3 Deep Neural Network for Fault Classification and Shapley Additive exPlanations approach for Explain the Model 63
3.3.1 Deep Neural Network For Faults Classification 64
3.3.2 Potential Ethical Impact of using AI-based Modeling 66
3.3.3 EXplainable Artificial Intelligence (XAI) Modeling 67
3.3.4 Overview of XAI Explanation Methods 69
3.4 Developed Framework 75
3.4.1 Sensor Data and Data Preparation 75
3.4.2 Data Analysis, Data Sampling, and Cross-Validation 76
3.4.3 Development of optimal multi-class classification model 76
3.4.4 Evaluation and Performance Metrics 77
3.4.5 XAI and Interpretation Blocks 79
3.5 Hydraulic System and Sensor Data Description 79
3.6 Results of Developed Framework 81
3.6.1 Descriptive analysis 81
3.6.2 Training Configurations and Selecting the Final Optimal DNN Classifier Model 84
3.6.3 Deep Neural Network Multi-class Classification Results 85
3.6.4 Deep SHapley Additive exPlanations Results 89
3.7 Discussion 92
3.8 Conclusions and Future Work 98

Abstract

This chapter presents a detailed framework for Condition Monitoring (CM) based on hydraulic systems and multi-sensor data. Nowadays, the CM technique is increasingly deployed to optimize quality and manufacturing processes. It is used as a decision-making support tool in maintenance operations or activities. In this environment, the diagnosis, prognosis, and monitoring of interconnected machines have become crucial issues for improving the cost-effectiveness of manufacturing industries. Some models are available to monitor or predict the degradation of elements within a hydraulic system, such as coolers, valves, internal pump leakage, or the condition of the hydraulic accumulator. In this case, we have focused on a data-driven approach, concentrating on the Deep Neural Networks (DNN) multi-class classification for unbalanced data adapted to predict the actual operating states of the system. Despite their performance, questions remain concerning the reliability of the DNN as a "black-box" model when used in complex applications, notably regarding the decision-making processes and the possible ethical, socioeconomic, and transparency impacts on stakeholders. Regarding the explanation approach, we have exploited the Deep SHapley Additive exPlanations (DeepSHAP) methodologies to provide reliable results and to explain the importance (weight) or role that each sensor plays and its contribution to the classifier algorithm's decision-making. The obtained framework based on two principal modules illustrates that the DNN classifier model when evaluated by Accuracy, F1-Score, Recall, and Precision metrics, are robust and perform efficiently. Finally, using the DeepSHAP technique explains the results of the developed model. It helps humans to understand, interpret and trust the model, with an associated increase in the support or the stimulation of Artificial Intelligence (AI) models applications on large-scale problems including industrial sectors.

Keywords: *Hydraulic system, Deep Neural Networks (DNN), Faults Classification, Sensors, Condition Monitoring (CM), Anomaly detection, Diagnosis, eXplainable Artificial Intelligence (XAI), DeepSHAP, Ethics*

3.1 Introduction

Hydraulic systems are important, due to their various applications in the industry, especially in the aeronautics, aerospace, and energy sectors [150]. These systems are generally composed of multiple interconnected machines, equipment, the Internet of Things (IoT), or sensors. Sometimes, their failures may occur randomly over time, if some countermeasures are not adopted in anticipation, these machines can fail to work properly or operate in an abnormal way. The maintenance strategies are very often periodical and do not really depend on the machine conditions [151]. This consideration is increasingly pertinent regarding security, reliability, performance requirements, and the provision of fairness. Moreover, in an increasingly competitive market, Industry 4.0 (I4.0) must respond to new challenges or opportunities related to profit, gain maximization, and mass or specialized production. These

challenges are driven by technical imperatives (e.g., minimization of machine downtime and maximization of a component lifetime) impacting economic issues such as the potential for production and maintenance cost reduction [11]. To meet these new demands, several techniques have been developed, including Condition Monitoring (CM) systems. In recent years, the CM approach has emerged as a key element in the management strategy and processes of industrial systems. Furthermore, emerging technologies such as smart and connected sensors, IoT, cloud computing, augmented reality, simulation tools, big data, and Artificial Intelligence (AI) techniques have contributed to the development and application of the CM system in the fourth Industrial Revolution [37], [152], [153]. More often, the CM technique has significantly improved its performance over other traditional maintenance methodologies including corrective, preventative, and time-based maintenance [154]. The most important task in performing CM concerns the monitoring of the machine components, sensors, and CM framework itself. Notably, the systems monitored can be subjected to potential failures (e.g., the alteration or breakage of the sensors) related to operating conditions or the effects of aging [155]–[157]. This approach tends to overlook the continuous changes based on the data collected online, resulting in an incapacity to detect abnormal values or to accurately predict the actual system's state. In addition, it is also necessary to be sure that the models developed do not interfere with the correct functioning of the other machines or components. For modeling condition monitoring or predictive maintenance tasks, different approaches have been developed [151] [158], [159]. These approaches can be classified into three categories: Knowledge-Based modeling [17]–[19], Physics-Based modeling [20]–[22] and Data-Driven or Data Science modeling [23]–[25]. Several papers have highlighted the performance of the data-driven technique, in particular, the robustness and performance of Deep Neural Networks (DNN) models in solving complex problems (e.g., fault detection, diagnosis), including non-linearity [160]–[163]. The DNNs models are a class of AI techniques named "black-box" models, which involve a combination of powerful learning algorithms with numerous parameters (thousands of parameters and layers) in the modeling space.

Despite their many applications and performances, the absence of transparency and explanation of the decision-making algorithms can constitute a major obstacle to their exploitation. Deep learning models are mostly presented as "black box" models. This is due to the fact that their internal predictive functions are not easy to explain or interpret by humans [164]. In addition to this, there are some questions regarding the large-scale application of the CM system based on AI techniques and the ethical impact such as transparency, trust, fairness, or privacy rules. Furthermore, Practitioners are interested in the explanatory paradigm of the complex AI algorithms [165]–[168]. The embedding of explanatory approaches to "black-box" modeling is an important requirement for the real-world experience of AI-based models. The main purpose of these methods is to address the different expectations, interests, goals, and needs of all stakeholders including citizens, regulators, governments, domain experts, or system developers [168]. Thus, the absence of an explanatory framework can have a significant impact on the fairness, ethics, transparency, responsibilities, and trust towards the outcomes of these models and their implementation on a large scale in organizations [169]. The key question must therefore be "*how to explain the decision-making process by the "black-box" models in a way that engenders faith in their reliability?*" In this context, it is necessary to improve both the accuracy, the performance, and the understanding of the decision logic or

mathematical rules used by these algorithms during the training stages. In addition, we must create a balance between explanation, interpretation, and accuracy, since explanation can assist us in the detection and correction of bias in the learning set (unbiased decision-making).

The main issue addressed in this chapter is the development of a detailed framework based on hydraulic systems and multi-sensor data. The framework has two principal modules; the first module addresses the prediction of the different degradation states of the hydraulic system components, whilst the second module concerns the explanation of the model's decision rules. This methodology provides an explanation or better understanding of the decisions of the model developed, while also characterizing their strengths and weaknesses in the decision-making process. We have focused on the DNN multi-class classification model for unbalanced classes, combined with the DeepSHAP approaches for the explanation of the model developed. To ensure robust performance and reduce the possibility of overfitting issues, we use cross-validation combined with a data resampling technique. Firstly, the developed framework is adapted to predict or classify the conditions of the hydraulic system, including the cooler, valve, internal pump leakage, or hydraulic accumulator. Secondly, the framework provides an explanation of the local and global importance (weight), or role played by each feature in the decision-making process. This facilitates the human comprehension of the mystery of Deep Learning (DL) algorithms and improves the interpretation of the results generated. It also encourages expert confidence in the use of AI techniques. According to our knowledge and the state of the art, very few studies are developing a hybrid framework for a hydraulic system. This framework combines two models: forecasting and an explanation XAI model. The first model is based on deep neural networks applied to unbalanced classes. The second model exploits the predicted classes by highlighting the advantages of the DeepSHAP approach. This model favors the implementation of "black box" models in industrial applications, including the conditional monitoring of a complex system in the context of Industry 4.0, from a methodological and operational point of view.

The rest of this study is organized as follows: Section 3.2 presents the literature review on the condition-based monitoring applied to hydraulic systems. Section 3.3 presents a robust DNN classification model, its characteristics, applications including the ethical impacts, and the reliability of the use in the condition monitoring systems. Furthermore, we introduce an overview of Explainable Artificial Intelligence approaches, and we describe their main concepts, definitions, and goals. We demonstrate that the exportation of the DeepSHAP technique for the DL model interpretation discriminates and explains more effectively the prediction performed by the DNN model. In addition, it is more adapted to human intuitions compared to other approaches. Section 3.4 shows a detailed description of the research methodology and the framework developed based mainly on two modules. The description of the hydraulic system including their available conditions and the data collected by the sensors are presented in section 3.5. Detailed results of the developed framework with in-depth predictive and explainable analysis are given in section 3.6. The major contributions and the discussion are presented in section 3.7. Finally, the conclusion, limitations, and future research are described in the last section of this chapter.

3.2 Literature Review of Condition Monitoring Applied to Hydraulic Systems

According to European Standards, maintenance is a combination of actions and management techniques that can be used to ensure the correct performance of the machine over time. We have three main stages for industrial maintenance: The first stage consists of identifying the faults and their characteristic if any. The second stage is the diagnosis phase which helps to identify the intrinsic location of the defects. The last phase named prognosis is based on the collected data and can be used to predict the operating states or Remaining Useful Life (RUL) of the machine. Hydraulic systems are also concerned by these maintenance strategies because such systems can be exposed to multiple conditions of fatigue, or failure. These causes are not necessarily recognized or identified. To explore them, the probabilities of occurrence, and to anticipate the deterioration levels in the hydraulic system condition, various models, or techniques have been proposed in recent literature [170]. Some papers have focused on monitoring the single component of the system, for example, the authors of [171], [172] have developed a hybrid prediction method based on the Empirical Mode Decomposition (EMD) and Support Vector Machine (SVM) to monitor the condition of a hydraulic transmission system with an axial piston pump. The authors of [173] focus on hybrid modeling based on adaptive neuro-fuzzy inference for pump state classification. Paper [170] investigates the behavior of a valve in a hydraulic system. In the following cases, the authors concentrate on several components of the system at the same time. The authors of [174] propose a statistical conditional maintenance approach based on Linear Discriminant Analysis (LDA) to predict typical defects related to some of the hydraulic components. Paper [175] exploits the hybrid approach as a combination of K-Nearest Neighbors (K-NN) and SVM models based on the multivariate time series for multi-class classification of the condition of the pump, valve, and accumulator. The authors of [176]] use Adaptive Linear Approximation (ALA) to extract local features in combination with the time-frequency wavelet decomposition analysis based on the signals. Papers [175], [176] propose a Kalman filter decomposition to perform the same analysis as the previous paper. In addition to this, paper [177] addresses a state-of-the-art overview of the conditional or predictive maintenance techniques for hydraulic cylinders. The authors highlight the major components of the system to monitor and their main potential failure factors such as water contamination, metallic debris, fluid fatigue, high temperatures, wear debris, sealing, and extrusion defects. In order to anticipate these defects, several approaches have been proposed [178], notably Electrical (magnetic) methods [178], [179], Optical methods [180], [181], Physical and chemical methods [182]. The Electrical approach, in particular, has been performed for the monitoring of the leak state in the cylinder. We can also mention other approaches such as the use of pressure sensors [183], [184], accelerometer sensors [185], [186], acoustic emission sensors [187] friction torque sensors [188], [189] and the Data-driven approaches [190], [191] which can be applied to drive many industrial systems. Most of these approaches are based on signal analysis, namely, on the time domain, frequency domain and frequency-time domain processed by Fourier transforms, Kalman, Hilbert filters, or wavelets decomposition. Paper [192] explains a numerical approach that aims to predict the RUL of an aviation hydraulic pump. In this case, the physical phenomena (wear debris characteristics) are simulated by the Monte Carlo sampling technique, and the RUL prediction

is based on the Partition-Integration method. This hybrid framework provides better results when compared to other simulation methods such as the Finite Element Method (FEM). Paper [174] uses the sensor data collected from the hydraulic system. Table 3.1 provides a summary of the models and strategies used in related work on this dataset. Especially, paper [174] uses the sensor data collected from the hydraulic system. These data are used to train some methods such as Support Vector Machine (SVM) Linear Discriminant Analysis (LDA) and Artificial Neural Networks (ANN). In addition, authors [193] used the method of feature selection or extraction. These features are used to train and validate several models like Logistic Regression (LR), Neural Network (NN), Decision Forest (DF), and SVM via benchmarks. To investigate the fatigue strength of hydraulic turbines and to reduce the probability of accidents or unplanned shutdowns (fatigue cracks in the impeller, or guidelines), the author of [194] proposes a useful and powerful intelligent maintenance system based on a statistical approach named Predictive Analytics System (PAS). It promotes the transition between scheduled maintenance and conditional maintenance by determining the optimal RUL of the system components. The authors of [195] introduce real-time monitoring of the condition of hydraulic oil, named the Impedance Detection System. This provides a prediction of premature equipment failures based on information that reflects the wear of the system. The approach has several advantages, it can be used for oil detection in the laboratory, or as a portable oil detection device for machine health monitoring.

In fact, we observe that most of the presented cited papers use Machine Learning (ML), Statistical, or Physics-based techniques. In addition to this, the data are frequently decomposed or approximated with the signal processing methods, or alternatively, the data are summarized as time series which requires several pre-processing and approximations operations. Despite their numerous applications and benefits, these approaches can also become less robust, unstable, and inefficient (overfitting problems) in the following cases: high-dimensional data, complex systems or machines, unbalanced classes, and missing and noisy data. Moreover, these models do not capture the causal relationship or the correlation between the descriptive and the predicted variables. Finally, the data decomposition techniques do not consider the local information combined with the temporal multivariate features. These limitations can significantly degrade the models and make them unusable. To overcome these challenges, the authors of [196] propose the Convolutional Neural Networks (CNN) to predict the condition of the hydraulic systems. The authors of [197] provide the data augmentation technique to improve the performance of CNN models when the data are not sufficiently representative. However, increasing the hidden layers have a degrading impact on network performance. Furthermore, the estimation of the model parameters may require a high-computing time or Graphics Processing Unit (GPU) memory.

In this chapter, we propose the multi-class classification method using the fully connected DNN model applied to the unbalanced classes which operate directly on data without any explicit selection or extraction of features. The aim is to explain the contribution of each explanatory variable by enhancing the work done by the authors of [196] which uses the CNNs models, and the local approach based on the Heat Map visualization technique. We have used DeepSHAP to explain the local and global importance of each sensor in learning states, to support all stakeholders in the decision-making process.

Table 3.1: Summary of work related to monitoring conditions applied to sensor data collected from the hydraulic system.

Reference	Feature Representation	Features	Classifier Model
Paper [174]	Engineered	Signal shape features + (slope of linear fit and position of maximum value) Distribution density (variance, skewness and kurtosis)	LDA, SVM, ANN
Paper [198]	Engineered	Distribution density: Median	RF and NB
Paper [193]	Engineered	Signal shape features + Distribution density (variance, skewness, and kurtosis)	ANN, SVM, LR, and DF
Paper [199]	Raw feature subset	Correlation and Redundancy-aware Feature Selection (CRFS)	LSTM-AE
paper [200]	Raw feature subset	PCA	XGBoost
Paper [196]	Raw sequence data	Encodings of the convolution layers	Fully connected CNN + Sensitivity Maps
Our proposed approach	Raw sequence data	Encodings of the Deep layers	Fully connected DNN + Explainable DeepSHAP

3.3 Deep Neural Network for Fault Classification and Shapley Additive exPlanations approach for Explain the Model

In this section, we present the multi-class classification approach based on a DNN model. Furthermore, we describe their architecture and internal structures such as the layer activation function, the number of neurons, the number of hidden layers, and the optimization function. Also, we discuss the potential ethical impacts of using AI-based modeling, notably the transparency, trust, fairness, and privacy rules. Moreover, we have discussed the paradigm for the most relevant explainable techniques of AI algorithms which are mainly based on the additive feature allocation methods. We analyze the contribution or the importance of each feature on the prediction of the DNN model. In other words, we show how each extracted

feature can influence the model. Thus allowing a local and global analysis of the data set and the problem to be solved. We can specify that the aim of this approach is not to evaluate the prediction quality or itself.

3.3.1 Deep Neural Network For Faults Classification

In this subsection, we focus on the data-driven module that consists of predicting the health conditions of a system by learning its behavior from historical data. We exploit the DNN model which is one of the most important architectures in AI-based modeling. This network has been developed and is used in many industrial applications [97], such as Computer Vision [201], Natural Language Processing [202], or Anomaly detection [203]. For our case study, we use a fully connected DNN model to classify the conditions of the hydraulic components such as the Cooler, Valve, Internal pump leak, Hydraulic accumulator condition, and Stable flag. Generally, this model improves the performance of traditional Machine Learning (ML) techniques with a better ability to generalize algorithms by mechanisms that rely on the regularity forms of the underlying learning functions. Moreover, DNN is a type of Artificial Neural Network (ANN) formed by several layers including an input and output layer and one or more fully connected hidden layers. The main objective of the DNN model consists of automatically extracting relevant features or patterns from collected data using networks with multiple hidden layers and nodes. Each layer of the network is composed of one or more interconnected artificial neurons. The information is processed in feed-forward mode (starting from the input layer and arriving at the output layers through the hidden layers). Thus, the expression of the input layer i is given by the vector $X = [x_1, x_2, ..., x_i]$. The hidden layer helps to extract features from the input layer x_i using the following equation:

$$h_j = \sum_{i=0}^{n} w_{ij} x_i + b_j \tag{3.1}$$

where $i = 1, 2, 3, .., n$, $j = 1, 2, 3, .., k$ is the number of hidden units. w_{ij} represents the kernel, b_j is the bias coefficients, x_i indicates a feature of previous layers. The inputs are multiplied by weights and the bias is added to the sum of the product obtained.

3.3.1.1 Layer Activation Functions

In the literature there are several families of activation functions classified into linear and non-linear functions [204]. To activate the hidden layers, we focus on the Rectified Linear Unit (ReLU) function [205]. This function is very often used for its simplicity, its ability to capture interactions, and its nonlinearity. Moreover, the application of the function of the hidden layers facilitates the gradient descent and promotes rapid training compared to other functions such as the sigmoid or hyperbolic tangent known as tanh, while being lighter and easier to calculate. It can learn reliably even when the number of layers increases. The monotonous function is just the max function (negative weights get flattened to 0), where the

relation is described by the following equation:

$$f_{ReLU}(x) = \max\{0, x\} \tag{3.2}$$

Concerning the DNN models, the activation function enables the set of neurons to be excited. The result is then converted into a signal indicating the state of neuron excitation. Thus, when we apply the ReLU function to activate the hidden layers, the result is presented by the following function:

$$h = f(h_j) \quad where \quad f(h_j) = ReLU(h_j) \tag{3.3}$$

However, to activate the output layers, we have two cases: when the variable to predict is binary, we use the sigmoid activation function. But if the variable has more than two classes (the model is a multi-class classifier), we apply the Softmax function (normalized exponential) which computes the probabilities of each subclass. The advantage of using these functions as the activation function for output layers is the fact that it is differentiable and compatible with the gradient algorithm. In addition, the Softmax function is a logistic function that takes as input a vector of j elements and as output a vector containing the normalized probabilities for each class whose sum is equal to 1. This function is defined by the following equation:

$$P(y = j|x_j) = \frac{e^{x^T w_j}}{\sum_{k=1}^{K} e^{x^T w_j}} \tag{3.4}$$

3.3.1.2 Loss Function and Optimizer

In this subsection, we introduce the general form of the loss function $\mathcal{L}$ for the multi-class classification model defined by the following equation:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} t_i log\ y_{i,c} \tag{3.5}$$

where N is the number of samples, C represents the number of classes, and $t_i \in (0, 1)$ with $t = 1$ for the correct class of the 1^{st} and 0 elsewhere. We use the Cross-Entropy function (CE) since the learning speed is faster than Mean Squared Errors (MSE). However, the DNN model's performance is often significantly affected by several hyper-parameters including the Learning Rate (LR). In order to minimize these errors, it is necessary to develop a neural network with an optimal architecture. In this context, we exploit the Adaptive Moment Estimation (Adam) method to compute this issue [206]. Thus, to adapt the learning rate, Adam's algorithm is based on the statistics moments including the first (mean) and second (variance) moments of the gradients. In addition to the CE function, we use Adam's algorithm

as an optimizer which is a process to find the optimal parameters through a learning stage by minimizing the $\mathcal{L}$ loss function.

3.3.1.3 Model Architecture

We have developed a fully connected DNN for Multi-Class Classification that solves tasks such as the prediction of degradation with multiple states in the hydraulic system's components. These components include the cooler, valve, internal pump leak, and the condition of the hydraulic accumulator. Figure 3.1 shows the architecture of the DNN model with one input layer, four hidden layers, and one output layer with nine nodes for each hidden layer.

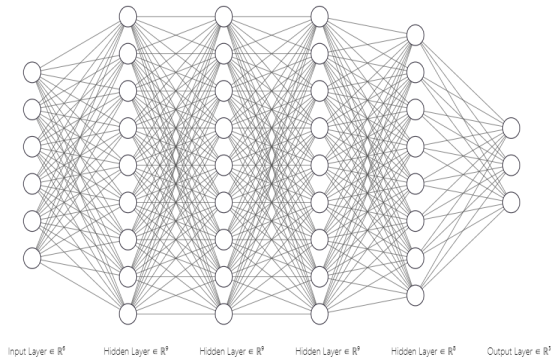


Figure 3.1: Architecture of the fully connected Deep Neural Network Multi-class classification model with four hidden layers, each layer contains nine neurons.

3.3.2 Potential Ethical Impact of using AI-based Modeling

We have highlighted that the AI-based models, particularly DNN, are important for industrial applications including the CM framework. Despite their numerous benefits, these models can lead to social-economic challenges, ethical impacts, or transparency issues concerning stakeholders. Thus, these issues can constitute a major obstacle to the large-scale application of AI technologies. The users are interested in the following aspects: *how to explain the decisions made by the algorithms in order to have confidence in them? Do intelligent systems have a level of responsibility in the moral or legal framework? How do we control the algorithm intelligence when they drift away from their target functions?* Research has been undertaken to highlight these challenging questions. Paper [207] shows that ethical principles and AI-based systems are connected at several levels, especially in ethics by design, ethics in design, and ethics for design. In particular, the authors [208], [209] argue that autonomous systems based on AI approaches are designed, deployed, and evaluated by humans. In addition, these systems are mainly based on theories, methodologies, or algorithms that incorporate principles or fundamental values such as transparency, legality, morality, and sociocultural fairness to ensure human flourishing and well-being. Furthermore, the authors [210], have

shown the necessity to ensure that the behavior of AI systems is beneficial to humanity. If the ethical principles are clearly and consistently defined for humans, the AI model itself will have the ability to make decisions based on these principles. In the efforts to make the system less biased in its decision-making, several approaches are possible. The first approach consists of generated scenarios that will help to challenge the system with virtual cases used in a simulated world, rather than in the real world. A second possibility is a formal approach that describes the automatic computation rules of a model and explains how a decision is or is not ethically acceptable. Moreover, to generalize the exploitation of the AI techniques, the stakeholders must have a common vision of their usage and objectives [151], [211].

3.3.3 EXplainable Artificial Intelligence (XAI) Modeling

3.3.3.1 Definition of some Concepts

In this subsection, we give some important concepts, definitions, and principles regarding XAI techniques. *Interpretability* is a passive characteristic of the models; it can determine if these models are meaningful or transparent for all users. Unlikely, *explainable* concept is an active characteristic, which means the capacity of the models to provide comprehensibly, and decision-making rules that can be analyzed by a human being. Moreover, *trustworthiness* is the confidence that an expert has in a model, i.e., the capacity of the model to produce the expected results [212]. However, *reliable* models do not completely conform to the requirements of explanation [213]. *Causality* is a principle to explain the causal relationship between the different input variables. *Transferability* is the understanding of the internal model relationships that facilitates implementation and usage while facilitating the transfer of knowledge to other problems. *Fairness* is the principle that any AI-based model must provide fair and equitable decisions without favoring any set of input data. In this sense, the XAI methods must respect the fairness constraints [214]. There are some additional associated principles to the XAI approaches, including bias, accessibility, and interactivity [215], [216].

3.3.3.2 Why Do We Use Explanations in AI?

To understand the use or importance of XAI approaches, we can provide the following questions; *How does a model work? What is driving decisions? Can we trust the model?* In this regard, the agnostic models for AI solutions are designed to be flexible and independent of the model parameters or intrinsic structure. The most common DL models provide results that are often incomprehensible or difficult to explain. Furthermore, we must be able to understand the decision-making rules generated by these models to guarantee trust and improve their performance. Therefore, the ability to interpret and understand the results of AI-based models is highly important. Figure 3.2 shows the comparison between the "black-box" AI models as a decision recommendation framework and explainable AI models as a set of explanation, feedback, and decision-making framework [217].

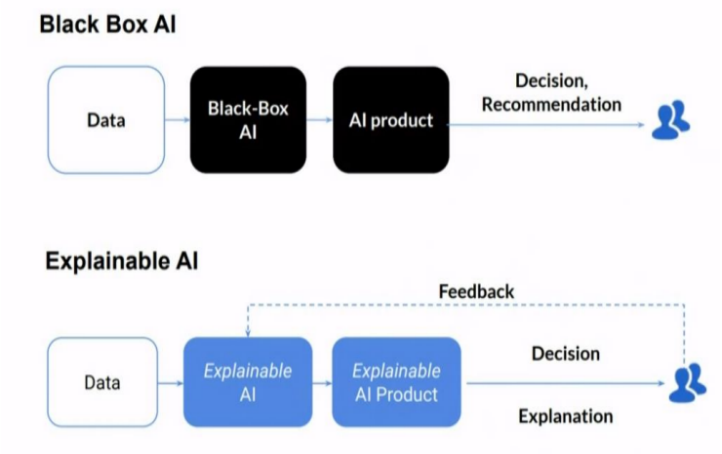


Figure 3.2: Framework of the "Black-Box" AI compared with the explainable XAI approaches.

3.3.3.3 What are XAI's Goals and their Main Target Audience?

We introduce the eXplainable Artificial Intelligence (XAI) methodologies as a sub-domain of AI techniques that provides a set of approaches or algorithms. The main objectives are to produce results that are easy to explain, and intuitive for all concerned, and that can also help to understand the internal functioning or structure of the AI models. In addition, this approach improves confidence among practitioners, insight into how the model can be improved, and understanding of the process being modeled. The XAI methods can be used in different contexts (See Figure 3.3). For example, data scientists use them to understand or to ameliorate the performance of the models, since they can help to improve and debug the models during the test and validation phase. In addition, manufacturers exploit it to check the robustness of the model and the impact on the production process. The explanation is therefore a helpful technique to provide transparency regarding how decisions are made and how they can potentially affect the users. Finally, these techniques address some concerns relating to the models, such as accuracy, robustness, bias, and transferability [165].

3.3.3.4 Levels of Transparency in XAI

In this subsection, we introduce the notion of transparency as the set of factors that facilitate the understanding and internal functioning of models including fidelity, accuracy, and generality. All elements of the decision-making process of the model must be able to be fully simulated by humans and an illustration is provided by models such as decision trees. A further property of the models is decomposability and algorithmic transparency; it is necessary to understand and analyze the set of procedures or mathematical rules followed by the model to generate their results. In this context, models such as DNN have a complex learning regime based on the optimization of a loss function, which is obtained through an approximation. Apart from this, their deep architectures do not necessarily meet the transparency rules.

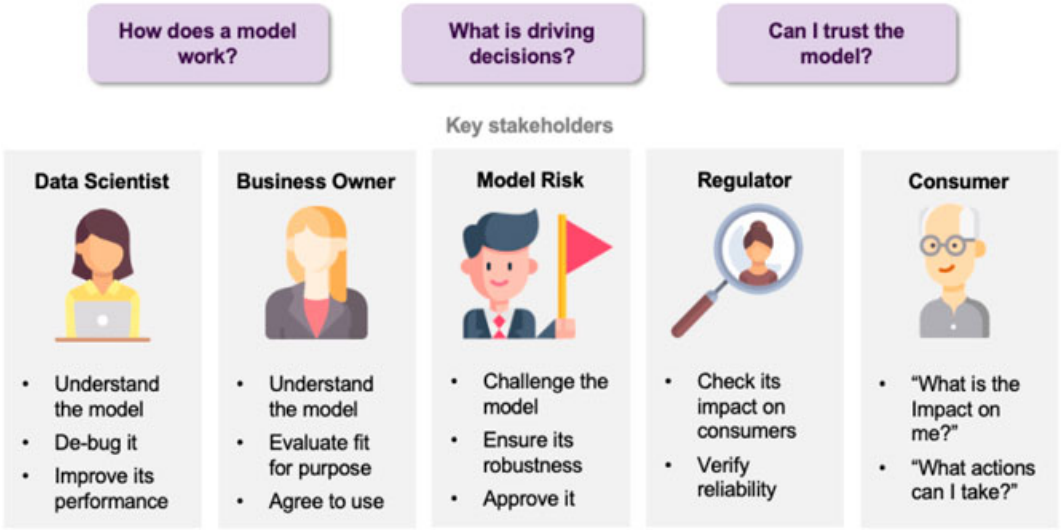


Figure 3.3: The interests of all stakeholders in the use and importance of the XAI approaches.

3.3.4 Overview of XAI Explanation Methods

The XAI explanation has three main family approaches: the local, global, and hybrid methodologies. The local explanation method expresses the contribution of each X input data instance in the decision-making to get a Y output in the training stage. This approach focuses on single input data to extract information about the explanation g by using the various characteristics of these data. The global XAI approaches provide insight into the set of decision-making rules applied by the model. In that case, we observe all features for the comprehensibility of the model. This approach is important for analyzing the global behavior of an AI model on all input variables. The last approach is the hybrid XAI explanation which combines the above-mentioned approaches. To address the problem of *How to explain the results of a Deep Learning algorithm?* we present the most popular explainable XAI methods developed in recent literature.

3.3.4.1 Activation Maximization

The activation maximization method is based on an input that maximizes the activation of a given hidden unit [218]. Its main purpose is to solve an optimization problem by maximizing the activation of a unit represented by the equation:

$$x^* = \underset{x \ s.t. \ \ ||x||=\rho}{\operatorname{argmax}} \ z_{ij} \ (\theta, x_i) \tag{3.6}$$

where θ, z_{ij} (θ, x_i) are respectively the parameter, and the activation of an individual unit i from the layer j. By fixing the parameter θ, we can obtain an explanation map. To resolve

the problem of explanation, we have several solutions that provide an overview of the features and their respective importance in each input layer. Firstly, the arithmetic mean of minimal values can be calculated. Secondly, the minimum value that maximizes the activation when the function converges is selected.

3.3.4.2 Layer-Wise Relevance Back-Propagation

The Layer-wise Relevance Back-Propagation (LRP) approach is developed by [219]. The authors of [220] have shown the importance of LRP approaches to explain the behavior of DL models. In addition, this method computes the most relevant scores on the input data characteristics by performing the decomposition of the obtained predictions. To explain the aim of this method, we consider a DNN model with an input x, a linear output y, and an activation output h, with a linear function $y_j = \sum_i w_{ij}x_i + b_j$ and $h_j = f(y_i)$. We note that the output neurons y_j are a function of the input neurons and the parameters of the fully connected layers. The relevance score $R(x)$ of the corresponding input variable x is given by the equation:

$$R(x) = \sum_j R_{i \leftarrow j} \quad with \quad R_{i \leftarrow j} = \sum_j R(h_j) \frac{x_i w_{ij}}{y_j + \epsilon |y_j|} \tag{3.7}$$

where $R(h_j)$ is the relevance of the activation output j and $R_{i \leftarrow j}$ is the Relevance or message of all nodes i that contribute to neuron j in the layer. In practice, if we apply this method in the CNN outcomes, the relevance score is back-propagated layer by layer. However, if the models are RNN, these scores are retro-propagated to the hidden states and the memory cell.

3.3.4.3 Saliency Map Visualization

Saliency Map Visualization (SMV) is introduced by [221], it is an explainable model using visualization to provide a gradient of the class output on a given input image. We can obtain an accurate summary of the input image y by viewing the positive gradients that have significant weight in the model output. Moreover, the visualization models allow us to find a class score function that approximates the score S_c defined by the following equation:

$$S_c(I) \approx w^T x + b \tag{3.8}$$

where x is the input image with a label class c which is a first order Taylor expansion, w represents the derivative of the score function S_c at a given point x_0 on the x input model and b is the bias coefficient of the model.

3.3.4.4 Deep Learning Important Features

Deep Learning Important FeaTures (DeepLIFT) is an additive feature attribution method that satisfies the local precision and recursively explains the results of the DNN model [222].

This method uses the linear composition method to linearize the non-linear elements of the "black-box" model [223]. Furthermore, DeepLIFT assigns a $C_{(\Delta_{x_i},\Delta_O)}$ value to the individual x_i input variables. These values describe the effects of a fixed input at a reference value relative to the original one. We introduce the function $x = h_x(x^{'}) \in (0,1)$ and we note that this method uses a "Summation-to-delta" property that is represented by the following equation:

$$\sum_{i=1}^{n} C_{(\Delta_{x_i},\Delta_O)} = \Delta_O \tag{3.9}$$

where $C_{(\Delta_{x_i},\Delta_O)}$ represent the contribution of the neuron x to the neuron y, $O = f(x)$ represents the output of the neurons, $\Delta_O = f(x) - f(r)$, $\Delta_x = x_i - r(i)$, and r is the reference input.

3.3.4.5 Local Interpretable Model-Agnostic Explanations

Local Interpretable Model-Agnostic Explanations (LIME) is an additive feature attribution, which aims to explain the prediction of a given AI-based model, by substituting it with a locally faithful explanation surrogate model [144]. LIME attempts to faithfully explain the predictions of the models by learning a locally interpretable model that is close to the prediction. Furthermore, this approach is a versatile explainer capable of processing various data types and models. Thus, the LIME algorithm is a local XAI approach that explains the prediction of a variable when analyzing its neighborhood. Let g be a class of possibly interpretable G family models such that $g \in G$, and $\Omega(g)$ represent a measure of the complexity of the interpretable model. The distance between two instances (x and z) around x is given by the measure $\pi_x(z)$. We introduce the function $\mathcal{L}(f, g, \pi_x)$ that is a faithfulness index of g approximating f in the locality as defined by $\pi(x)$. Thus, the explanation $\xi(x)$ for an input x is given by the following equation:

$$\xi(x) = \underset{g \in G}{\operatorname{argmax}}\, \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{3.10}$$

In the practical case, LIME optimizes only the loss part function using Penalized Linear Regression (PLR) technique. To determine the appropriate model complexity, it is possible to select the maximum number of features that the model can use during a training stage. Being an agnostic model, it is applicable to any "black-box" model. Furthermore, it has been used in several analyses including text and image processing. LIME allows a better qualitative understanding of the influence of each input variable on the output predictions. However, this approach has some limitations related to its local explicability, it does not generalize local interpretability results to a global level. Moreover, LIME provides unsatisfactory results in the case of tabular data and when the explanatory variables are of a continuous or categorical type. Several LIMEs approaches have been developed to address these issues, such as the Sound-LIME (SLIME) approach [224], which helps to identify the temporal-frequential regions having a major impact on the decision-making model. In this way, temporal descriptions

or decisions that are less intuitive are transformed into more insightful spectral information. In addition, the explanations generated provide a deeper understanding of the behavior of the classifier in order to determine an unreliable and non-generalizable. Kullback-Leibler LIME (KL-LIME) is a combination of LIME and the principle of Bayesian projection predictive feature selection methods, for explaining the model CNN predictions [225]. Furthermore, KL-LIME represents the compromise between the faithfulness of the explanation and its complexity. The quadratic LIME (QLIME) approach rescales LIME's linary relationships to a quadratic relationship, with improvements in the accuracy of feature interpretations and minimizes the Root Mean Square Error (RMSE) of the predictive model [226]. Finally, Modified Perturbed Sampling for LIME (MPS-LIME) approach is based on a perturbed sample method where instances are being generated from a uniform distribution, which ignores complex correlations between different features [227]. The MPS-LIMEs can address the limitations of the classical LIME sampling operations.

3.3.4.6 SHapley Additive ExPlanations

Shapley Additive exPlanations (SHAP) is based on the Shapley values, combining both the game theory and the local explanation methods [228]. The SHAP approach is more efficient since it allows a complete explanation. Furthermore, it is possible to compute the global measure by means of aggregating the local feature importance for each observation. In addition to these mathematical properties, the authors of [228] prove that SHAP has three further properties including local accuracy, the loss of constraint features, and consistency. To predict an event, we use the Shapley values to determine the importance of local features for each observation. We can approximate Shapley values as additive feature assignments. These values determine the effects of each variable characteristic of a predictive model. Moreover, these values show how the set of features impacts the predicted outputs of a given "black-box" model. This explanation approach also provides the ability to determine and visualize the attribute characteristics defined by the "force" or Shapley values. Let $z^{'}$ be the "Coalition vector" and g the explanation model represented by the equation of the additive feature attribution methods:

$$g(z^{'}) = \phi_0 + \sum_{j=1}^{M} z_j^{'} \phi_j \tag{3.11}$$

where ϕ_0 is the base value, the "coalition vector" $z \in (0,1)^M$, with $z_j^{'} = 1$ if the value of the corresponding characteristic is "present" and $z_j^{'} = 0$ when the value is "absent". So, M represents the maximum coalition size or the number of simplified input features, $\phi_j \in \mathbb{R}$ is the feature attribution for j, and $g(z^{'})$ is the sum of the contributions of the biases and the individual characteristics or the value predicted by the model for this instance. To determine the Shapley value or the feature importance of the model predictions, we exploit the cooperative game theory results. The classic Shapley value estimation has more forms such as the Shapley regression values, and the Shapley sampling values [228]. Shapley regression values ϕ_j are usually applied for linear models in presence of multicollinearity. These values represent a unique solution to the additive agnostic interpretability problem of the "black-box" model. Furthermore, the values satisfied some properties including accuracy, missingness, and

consistency [228], [229]. The expression of ϕ_j is given by the following equation:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}}^{n} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \Big[f_{SU\{j\}}(x_{SU\{j\}}) - f_S(x_s) \Big] \tag{3.12}$$

where $S \subseteq F$ is the subsets features F represents the entire set of features, and $SU\{j\}$ is the union of the subset S for feature j. The Shapley regression values are usually used as feature attributions, these values can be considered as a weighted average of all the possible differences. Shapley's sample value is based on re-training a model on all subset's features [228]. It allocates to each feature an importance value or an effect of the inclusion of this feature on the predicted model. Subsequently, both models ($f_{SU\{j\}}$, and f_S) are then respectively trained with the present and withheld features. Then the results of both models are compared on the current input $f_{SU\{j\}}(x_{SU\{j\}}) - f_s(x_S)$ where x_S is the value of the input features in the set S. Shapley sampling values consist of applying the sampling approximations to equation 3.12. This estimation method combines a sample of the training data set to approximate the effect of removing variables from the model. In addition, the approach has several advantages, it is unnecessary to re-train the model, thus significantly reducing the number of errors and the computational time. Therefore, we use Shapley sampling values because it applies to all models and has more advantages than Shapley regression values. To analyze the aggregated or global importance features that are explained by the DeepSHAP method, we use the Shapley mean absolute value described by the following equation:

$$\frac{1}{N} V_j = \sum_{i=1}^{n} |\phi_j^{(i)}| \tag{3.13}$$

The SHAP approach is an additive feature attribution method, thus, the prediction can be expressed as the sum of the different effects of the explanatory variables. In the presence of massive data, this approach has a high memory cost and usually requires an exponential time to compute all Shapley values which can generate some delays in the estimation of the predictive model and the real-time interpretation process. To avoid these inconveniences, several alternative approaches have been proposed [228], [230]–[232]. The first approach is the TreeSHAP, which is a specific implementation of Shapley's explanations based on the Decision Tree (DT) model theory [230]. This approach provides a consistent and complete explanation of the decision-making rules made by the DL models or ensemble algorithms including the Random Forest (RF). However, for certain models such as Extreme Gradient Boosting (XGBoost), the TreeShap can be locally inaccurate and suffers from irregularities due to the instability of the DT models.

The second methodology is named KernelSHAP, which is a linear combination of LIME and Shapley values. KernelSHAP can be applied to some ML and DL models. Algorithm 1 shows the pseudo-code of the method, it has mainly five main steps [224], [233]. The purpose of this algorithm consists of performing an additive feature attribution through random sampling of the coalition vectors, by extracting features from the input data and making a linearization of the model influence using SHAP kernels. However, the accurate estimation of the Shapley

values can require significant computational time since the method needs a combination of all values of each variable in the considered model on the whole data set. This algorithm is composed of several functions. The function *SampleByRemovingFeature(x)* aims to compute the features; it can be decomposed in two stages. The first stage consists of applying sampling approximations to equation 3.12 and the second one is to approximate the effect of removing the variable on the model by including the training data set samples. Thus, the feature $h_x \in (0, 1)$ can be mapped to the original input space, where $h_x = 1$ means that the input is included in the model, otherwise, it is excluded from the model. Each feature value extracted (simplified input mapping h_x) is reshaped to a similar input size, and then saved in the list z_k. Furthermore, we used the simplified input mapping h_x to compute the function $g(z_k)$. To compute the associated weights W_x of the coalition vectors, we apply the function $SHAP$ (g, z_k, y_k) that aims to build the local explanation model and to obtain the Shapley values. The function is represented by the given relation $g_x(z_k) = g(h_x(z_k))$ where S represents the set of non-zero indexes in z and z_S is the missing value for the features that are not included in the set S. Since most models are not able to support arbitrary patterns of missing input values, we use an approximation $g(z_S)$ with $E[g(z)|z_S]$ to compute the additive feature contribution.

$$g_x(z_k) = g(h_x(z_k)) = E[g(z)|z_S] \tag{3.14}$$
$$= E_{z_{\bar{s}}|z_S}[g(z)] \tag{3.15}$$
$$\approx E_{z_{\bar{s}}}[g(z)] \tag{3.16}$$
$$\approx g([z_S, E[z_{\bar{s}}]]) \tag{3.17}$$

where $\bar{s}$ is the set of features not in S, and $E[g(z)]$ represents the base value. The equalities of the equations (14 and 15) represent respectively the SHAP explanation model simplified input mapping and expectation over $z_{\bar{s}}|z_S$. The approximation (16 and 17) assumes notable feature independence and model linearity. In the practical case, The *SHAP(.)* computes the Shapley value by using the average of ϕ_j. The result of this step is a two-dimensional list W_x, where each row is the most important Shapley value for one feature from the list. The function $Model(W_x).fit()$ is used to explain the model's predictions restricted to the feature space S applied to x_s. Finally, the function *Model.coefficients()* builds and returns all the coefficients of the explained model.

Algorithm 1: KernelSHAP approach for the Local Explanations

Input Classifier g, input sample x
Output Explainable coefficients from the model

1: $z_k \leftarrow$ SampleByRemovingFeature(x)
2: $z_k \leftarrow h_x(z_k)$ ▷ h_x is a feature transformation to reshape to x
3: $y_k \leftarrow g(z_k)$
4: $W_x \leftarrow SHAP(g, z_k, y_k)$
5: $Model(W_x).fit()$
6: Return Model.coefficients()

The last approach is the DeepSHAP which is considered a mixture of DeepLIFT and Shapley values. The authors of [228] demonstrate that under the linearity assumptions on

the DL parameters and the independence between the input variables, the Shapley values are close to those of the DeepLIFT. Moreover, the local precision and the property verification (e.g., presence or absence of an object) of DeepLIFT combined with the coherence and the desirable properties (e.g., efficiency or symmetry) of Shapley values motivate the adaptation of the DeepLIFT to approximate the Shapley values, resulting in the DeepSHAP model. This explainable approach has several advantages including the local and global explanation of the contributions of each input variable. It also exploits the DL features to improve computational performances and extract deep information.

3.4 Developed Framework

Figure 3.4 illustrates the proposed framework which has mainly two purposes and three stages. The first aim is to perform the data pre-processing (first stage), subsequently, the data are used for training and validating the DNN model (second stage). The resulting model is a DNN optimal classification model which predicts the state of the hydraulic system. The second goal of the approach consists of using the output of the resulting model as input for the explainable XAI technique (local or global). This explainable model named DeepSHAP is performed to explain, visualize, and identify the most important feature's contribution of each sensor in the model decision-making rules. In this context, the combination of these two models provides robust high-performance for the CM strategies, enabling the classification results to be evaluated, interpreted, or explained. Moreover, we can note that these advantages encourage the use of AI-based modeling and CMs strategies in the autonomous systems industries. In the following sub-section, we give a detailed description of each block of the proposed methodology.

3.4.1 Sensor Data and Data Preparation

The first two blocks of the diagram represent respectively the database and the set of data preparation. A detailed description of the used data sets which describe the hydraulic system including their operating conditions is illustrated in section 3.5. These structured data contain both categorical (see table 3.3) and numerical data (see table 3.2). To perform the model, we need to perform several pre-processing operations on the data and subsequently used it as inputs to train and validate the model. Thus, the classic data pre-processing is the formatting of any raw data, the data normalization, and the One-Hot-Encoding applied to change the target categorical to numerical data.

(a) Min-Max Normalization: The normalization operator is applied to numerical data; this technique is required when the features have widely differing ranges. the operation limits the effect of the size of each feature during the learning stage. In our case study, we apply the Min-Max normalization technique defined by the following equation:

$$X_{Norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad where \;\; X_{Norm} \in (0, 1) \tag{3.18}$$

(b) One-Hot-Encoding: The One-Hot-Encoding technique is a simple and efficient operation usually applied to categorical data pre-processing in AI tasks. This processing operator consists of converting the values of a multi-state variable on 1 bit (number of the state assumed by the variable) and keeping other elements to 0. For example, the cooler conditions have three distinct states, using this operator, we can convert the cooler states as follows: 'Close to total failure' is encoded as $(1, 0, 0)$, 'Reduced efficiency' is encoded as $(0, 1, 0)$, and 'Full efficiency' is encoded $(0, 0, 1)$. Furthermore, this strategy is also applied to all states or classes of the target variables including valve, internal pump leakage, hydraulic accumulator conditions, and stale flag.

3.4.2 Data Analysis, Data Sampling, and Cross-Validation

The third block shows all the descriptive analyses including the summary statistics performed on the normalized and non-normalized data. These analyses provide insights and help to capture possible trends, patterns, or anomalies in the data. We note that some output variables (classes) are unbalanced and in addition, there are not any missing values. The main results are illustrated in sub-section 3.6.1.

3.4.3 Development of optimal multi-class classification model

The block named DNN classifier model consists of a fine-tuning process and a prediction principle. The prediction is based on an algorithm that trains a historical dataset and applies to new data when estimating the likelihood of a specific outcome. In this setting, the prediction concerns the conditions of the hydraulic system. To achieve this goal, we develop a fully connected neural network that considers as input several explicative variables and as output, we have the predicted variables. We note that these variables have been transformed in different processes (see block 2). During the training process of the DNN model, the parameters of the layers are randomly initialized, via the re-sampling procedure combined with the stratified k-fold cross-validation technique. This process is then repeated m times; thus, the resulting model is considered the best model of cross-validation. The comparison between all the generated models is performed by choosing the highest precision or F1 score values. Regarding the data sampling and division, we applied the Cross-Validation coupled with the re-sampling technique (see the fourth block of diagram 3.4). Usually, to train any AI-based model, it is possible to split the data randomly into a training set (70%) and a test set (30%). The first set is used for training, while the second one helps to validate the model. However, this approach can lead to overfitting problems, since not all the samples are tested equally. To avoid this problem by assuring a generalized DNN, we use stratified K-fold cross-validation

algorithms with k divisions and m repetitions which integrate the sampling techniques. This operation preserves the distribution of each fold to ensure optimal training of the data set.

(a) If the best model obtained in the previous step does not satisfy the evaluation criterion then this model is not performing. We, therefore, return to the training step. So, we repeat the fine-tuning process by using the same data and by exploiting the weights of this model to initialize a new model. Like the model training stage, the backpropagation algorithm is used to update the gradients of all outputs. In other words, the fine-tuning consists of adjusting the hyperparameters, by evaluating the model with several learning rates and a number of epochs. This phase is then repeated as long as the model has not satisfied the performance requirement.

(b) However, if the obtained model performs efficiently, then it will be considered a potential candidate optimal model. To validate this model, we exploit new data which did not participate in the model training. A detailed description of the training configurations and selecting the optimal DNN Classifier model is described in the subsection (3.6.2).

3.4.4 Evaluation and Performance Metrics

To evaluate the performance of the developed DNN model, we use metrics including Accuracy, F1-Score, Precision, and Recall. The Accuracy metric measures the proportion between the 'True' predicted values and the total number of sample data. However, this metric ignores 'False' predicted values. Therefore, to address this issue we can perform the precision or the recall metrics. The Precision metric provides a measure of the number of correct classifications versus the number of incorrect classifications states. The Recall metric measures the number of 'True' classifications with the number of 'False' entries. Often it is necessary to investigate additional metrics to confirm that a given model is valid, in this case, we can use F1-Score. This metric attempts to stabilize the precision and recall metrics by performing the harmonic mean between them. The equations for each metric are described as follows:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ Number\ of\ Predictions} \tag{3.19a}$$

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{3.19b}$$

$$Recall = \frac{True\ Positives}{True\ Positives\ +\ False\ Negatives} \tag{3.19c}$$

$$F1\ Score = 2\ \times \frac{Precision \times Recall}{Precision\ +\ Recall} \tag{3.19d}$$

where the True Positives (TP) represents the number of samples classified correctly, False Positives (FP) is the number of samples classified incorrectly, True Negatives (TN) is the number of samples classified correctly as a normal class and False Negatives (FN) is the number of samples classified correctly as a normal class for each target variable. During the

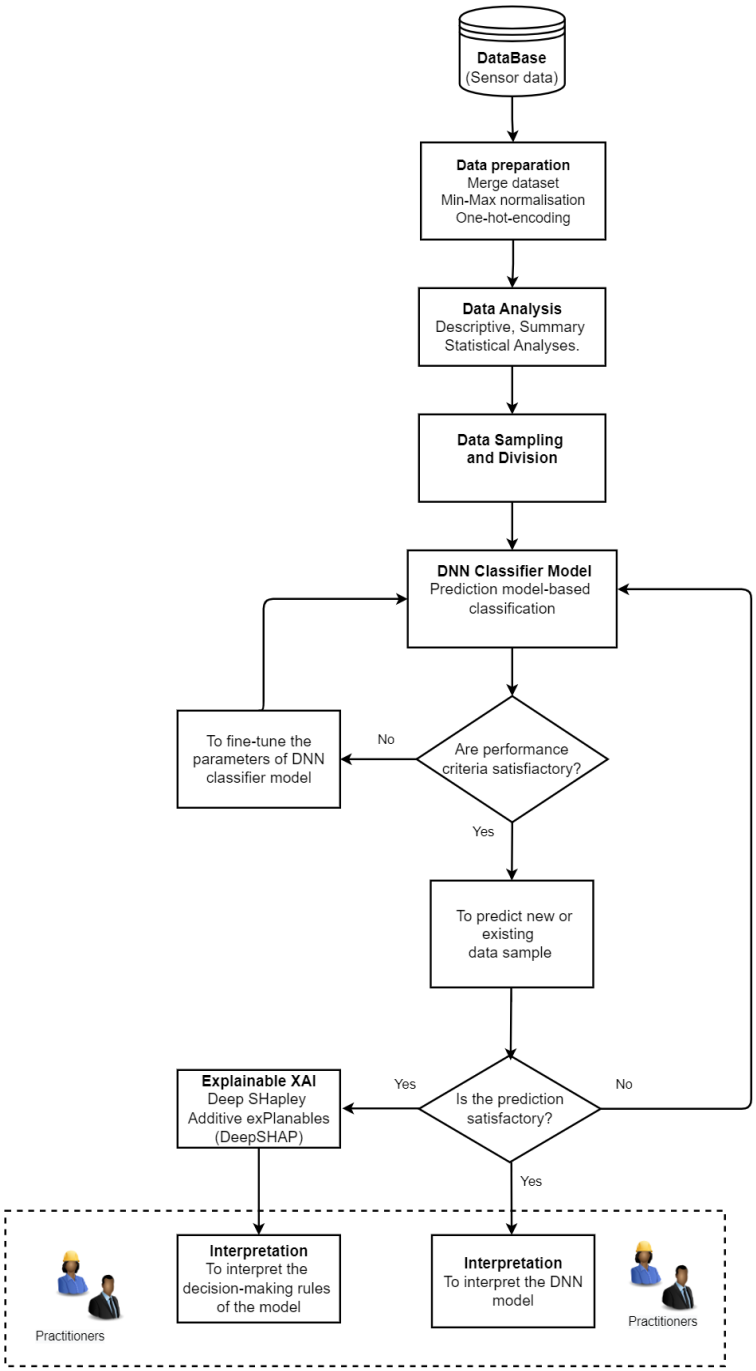


Figure 3.4: Diagram of the proposed framework. The workflow has two principal modules: (a) The DNN multi-class classification module aims to predict the different conditions for failure of the hydraulic system components. (b) The explainable DeepSHAP module provides insight into the decision-making process of the classifier model.

fine-tuning process, the best predictive DNN classifier model can be selected according to the highest Accuracy and F1-Score value metric.

3.4.5 XAI and Interpretation Blocks

The last two blocks (XAI model and Interpretation) deal with the second objective which consists in explaining the outcomes of the developed multi-classification model and supporting humans to understand and interpret the mechanisms of this model. Thus, the XAI like DeepShap takes as input the predicted values of the optimal DNN classifier model and explains the output of the used DL model. The main idea is to evaluate the average impact of a variable for all possible combinations of variables. Moreover, by averaging the absolute values of the Shap values for each variable, we can trace the overall contribution or importance of the variables (see section 3.6.4)

3.5 Hydraulic System and Sensor Data Description

Hydraulic systems have many industrial applications and interests [234], [235] such as power energy production, the ability to drive heavy charges, or multiple machines. However, these systems can be subjected to several failure risks (e.g., due to high pressure or leaks that can impact their efficiency). Therefore, to anticipate these anomalies, it is crucial to control or monitor the operating conditions of the system. The monitoring operation can intervene at several levels in different components such as the valve, internal pumps, cooler or hydraulic accumulators [236]. Figure 3.5 shows the investigate hydraulic system [174], [237] and the sensor data are available at the UCI Machine Learning Repository website [1]. These data were collected through an experimental study of the hydraulic test rig [236], the data set is composed of $43,680$ features (Distributions: $1Hz \rightarrow 8 \times 60 = 480$, $10Hz \rightarrow 2 \times 600 = 1200$ and $100Hz \rightarrow 7 \times 6000 = 42000$). In addition, the frequency acquisition of 17 sensors (physical and virtual types) installed on the circuit is around 60 seconds. The multisensor dataset are described in tables 3.2 and 3.3. In particular, table 3.2 shows the description of quantitative data such as pressure, speed, flow, power, temperature, or vibration measurements. Table 3.3 describes the target variables and the states concerning each component deterioration, such as pressure leak in the accumulator, internal of the pump, delay in the switching of the valve, reduction of the cooling efficiency, and volume flows. We can note that the target variables (predicted variable) such as Cooler, Valve, Internal pump leakage and Hydraulic accumulator conditions have more than two states and the Stable flag has a binary state.

[1] https://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems

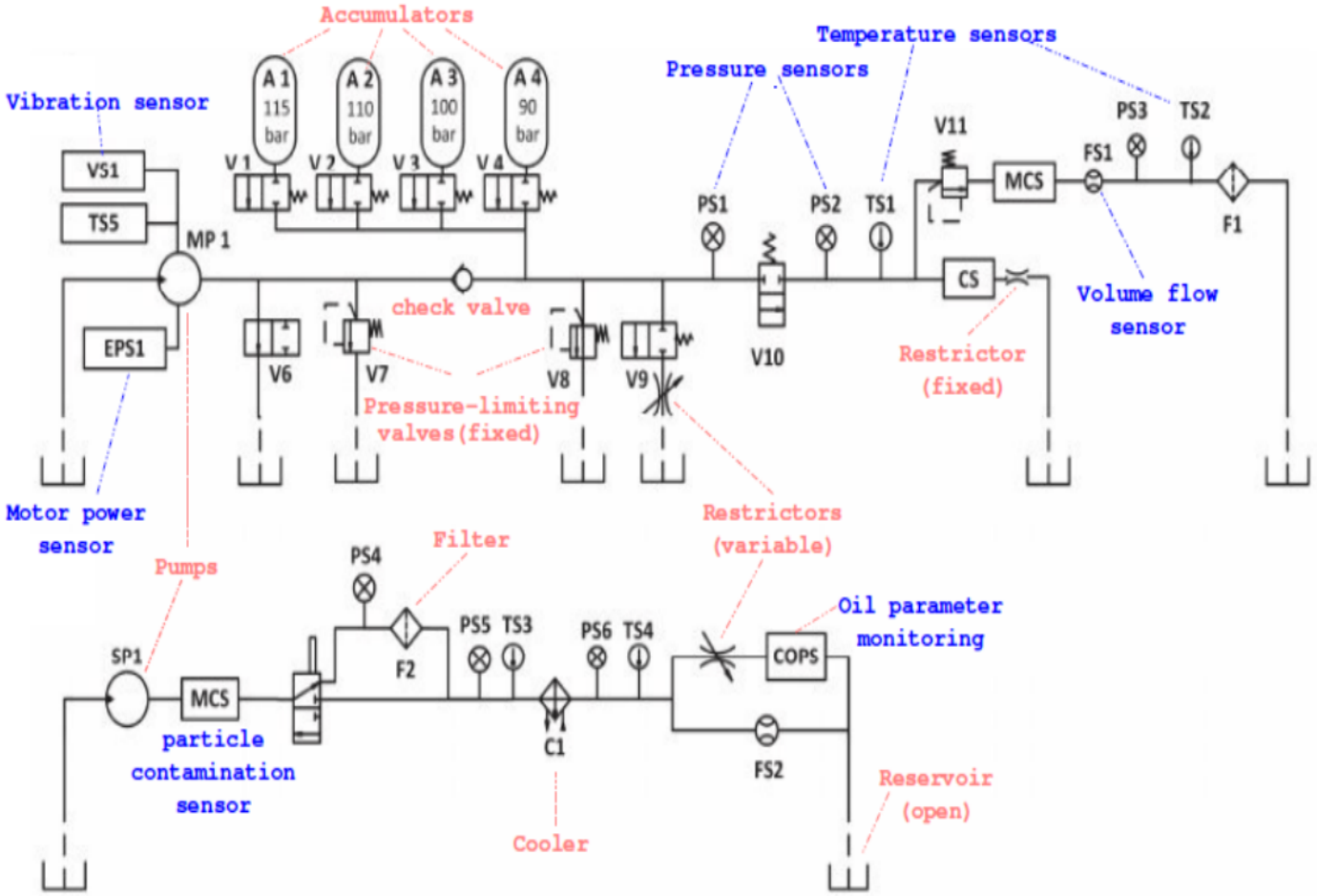


Figure 3.5: Diagram of the hydraulic test bench: The test rig is composed of two main circuits, namely the primary working circuit to control the load and the secondary cooling-filtration circuit, which are connected by the oil tank. An upper circuit is the primary working circuit, it is connected to the lower circuit which provides cooling and filtration via the oil tank; the sensors are highlighted in blue [198], [236].

Table 3.2: Physical data (quantitative data) collected by the sensor.

Number	Sensors	Physical Quantity	Sampling Rate	Unit
#1	PS1	Pressure	100	Bar
#2	PS2	Pressure	100	Bar
#3	PS3	Pressure	100	Bar
#4	PS4	Pressure	100	Bar
#5	PS5	Pressure	100	Bar
#6	PS6	Pressure	100	Bar
#7	TS1	Temperature	1	$°C$
#8	TS2	Temperature	1	$°C$
#9	TS3	Temperature	1	$°C$
#10	TS4	Temperature	1	$°C$
#11	VS1	Vibration	1	mm/s
#12	SE	Efficiency factor	1	%
#13	EPS1	Motor power	100	W
#14	FS1	Volume flow	10	L/min
#15	FS2	Volume flow	10	L/min
#16	CE	Cooling efficiency	1	%
#17	CP	Cooling power	1	W

3.6 Results of Developed Framework

3.6.1 Descriptive analysis

The analysis was conducted with Jupyter notebook, and we used Python modules such as TensorFlow, Scikit-Learn, Keras, and Shap. As with any AI model analysis, some data pre-processing operations were necessary and included formatting, Min-Max normalization, and the One-Hot-Encoding technique applied to the target variables. The histogram represented by figure 3.6 shows the frequency distributions of the 17 features contained in the data set. These variables are distributed between $(-1, 1)$, so we can observe that the explanatory variables do not fit with any distribution including the Gaussian distribution. Moreover, it is impossible to deduce any information about the existence of the outliers. To address the automatic classification problem in Deep Learning (DL), it is important to perform an analysis of the unbalanced classes. We observe that some sub-classes of the target variables are unbalanced, which may be due to the scarcity of the collected data. These sub-classes reflect the various cases of correct functioning or failure of the system. The cooler condition subclass is perfectly balanced (33%). However, the valve conditions sub-classes are unbalanced since the state 100 (Optimal switching behavior) is unbalanced compared with the remaining three states. In addition, for the condition of the hydraulic accumulator, all the sub-classes are

Table 3.3: Monitored parameters: Categorical data of the hydraulic test bench. Each variable represents a system operating multi-state.

Condition Variable	State	Cases Number	Target Value
Cooler Condition (%)	- Close to total failure	732	3
	- Reduced efficiency	732	20
	- Full efficiency	741	100
Valve Condition (%)	- Optimal switching behavior	1125	100
	- Small lag	360	90
	- Severe lag	360	800
	- Close to total failure	360	73
Internal Pump Leakage	- No leakage	1221	0
	- Weak leakage	492	1
	- Severe leakage	492	2
Hydraulic Accumulator	- Optimal pressure	599	130
	- Slightly reduced pressure	399	115
	-Severely reduced pressure	399	100
	- Close to total failure	808	90
Stable Flag	- Conditions were stable	1449	0
	- Static conditions might not have been reached yet	56	1

unbalanced.

In this study, we will show that the model obtained performs well and is more robust when dealing with the effects of unbalanced classes. Moreover, the DNN does not require a high computing time to generate the prediction results. The distribution of each target or predicted variable is the following:

1. Cooler Condition
    - Class=3, Count=732, Percentage=33.19%
    - Class=100, Count=741, Percentage=33.60%
    - Class=20, Count=732, Percentage=33.19%
2. Valve Condition
    - Class=100, Count=1125, Percentage=51.02%
    - Class=73, Count=360, Percentage=16.32%
    - Class=80, Count=360, Percentage=16.32%

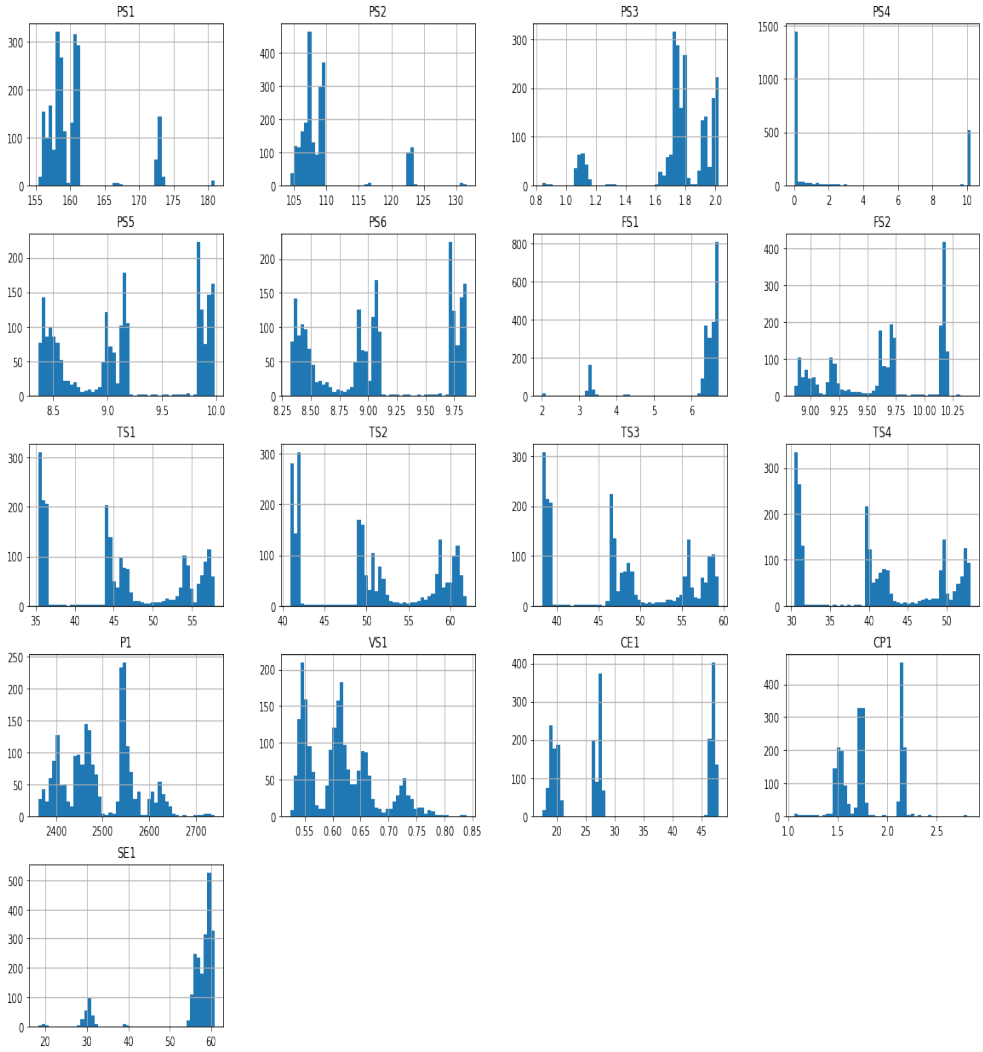


Figure 3.6: Frequency distribution of the sensors data: The histogram represents an average of each data cycle.

- Class=90, Count=360, Percentage=16.32%

3. Internal Pump Leak

   - Class=0, Count=1221, Percentage=55.37%
   - Class=2, Count=492, Percentage=22.31%
   - Class=1, Count=492, Percentage=22.31%

4. Hydraulic Accumulator

   - Class=130, Count=599, Percentage=27.16%
   - Class=115, Count=399, Percentage=18.09%

- Class=100, Count=399, Percentage=18.09%
- Class=90, Count=808, Percentage=36.64%

5. Stable Flag

   - Class=1, Count=756, Percentage=34.28%
   - Class=0, Count=1449, Percentage=65.71%

3.6.2 Training Configurations and Selecting the Final Optimal DNN Classifier Model

Feed-forward network development is driven by the selection of fine-tuned hyperparameters, especially the number of hidden layers, neurons per layer, epoch, learning rate, optimization, and layer activation function. In this case, we use Adam optimizer as the optimization function, ReLU for the activation layer, and Softmax for the hidden layer activation function. To improve the model performance, several steps are necessary such as training, optimization, fine-tuning, and validation of the model. Regarding the network training through stacked layers, the parameters of those layers including the output layers are randomly initialized. During this stage, we use re-sampling combined with the stratified and repeated k-fold cross-validation technique to train and challenge the DNN model since some of the target variables have slightly unbalanced classes. To obtain the validation data, we split the data set into k-folds ($k = 5, 10, 15, 20, 25,$ and 30). This technique helps to avoid the overfitting effects and $(k-1)$ folds are exploited to train the model and the remaining data are used to test the model. This process is then repeated m times where $m = 4$, and the obtained result is considered the result of cross-validation. The aim of the fine-tuning procedure of the entire network is to modify the weights of a trained neural network. We, therefore, exploit these weights to initialize a newly trained model using the same data. As for the training case, this process uses the back-propagation algorithm which updates the gradients of all layers (from the lowest to the highest). To build a powerful model, we also adjust the hyperparameters and thus evaluate the model with several learning rates ($1e^0, 1e^{-1}, 1e^{-2}, 1e^{-3}, 1e^{-4}$ and $1e^{-5}$) and epochs number ($50, 150, 200,$ and 250). Furthermore, we exploit a lower learning rate than that used during the model training, and we repeat the experiment several times. The comparison between several models generated is performed by choosing the highest Accuracy or F1 Score values (see Table 3.4). In summary, the approach consists in training the DNN classifier, then optimizing it thanks to the Adam optimizer, fine-tuning it, and finally evaluating the proposed model by using the cross-validation technique.

The proposed optimal Fully Connected Neural Network (FCNN) has the following architecture and parameters.

- Input layer shape: (2205×17)
- Number of hidden layers: 9
- Number of nodes for each hidden layer: 25×8

Table 3.4: Table shows the performance of the DNN classifier model for Cooler conditions. We use several values of learning rate to train the DNN. The best model selection is chosen according to the highest Accuracy value.

Values of lr	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0.1	47.89	52.26	51.69	50.65
0.01	99.87	99.84	99.84	99.84
0.001	99.87	99.39	99.39	99.38
0.0001	99.87	99.69	99.69	99.69
0.00001	82.76	70.68	70.90	71.75

- Learning rate: 0.01
- Number of learning iterations: 250
- Initial learning weight diameter: 0.1
- Momentum: 0.01
- Type of normalizer: Min-Max Normalizer
- Metric: Accuracy
- Optimization algorithm: Adam
- Loss function: CrossEntropy
- Training algorithm: Back-propagation
- Activation functions: ReLU and Softmax
- Output layer shape: $(2205 \times n)$, where n represents the number of the target variables' sub-classes. In fact, we have n equal to three for the cooler condition and internal pump leakage variable. Regarding the valve conditions and hydraulic accumulator variable, the number of states is four, and the unique variable with two states is the Stable flag.

The best parameters for training the performing DNN classifier model are as follows: The number of k-Folds is 20, the random state is 10, the batch size is 16 and the number of epochs is 250. Consequently, the final model's validated network structure and the number of hyperparameters are presented in table 3.5. This table shows the best-developed model summary, including the shape of each output layer, their weights, and the total number of parameters $(5,535)$ that must be estimated when training the DNN classifier model.

3.6.3 Deep Neural Network Multi-class Classification Results

In this subsection, we present the major results of DNN models including their performance metrics and Misclassification errors. The fine-tuning of the hyper-parameters model provides optimal predicted results. To test and validate the forecast results, we consider the best DNN multi-class classification model obtained through the fine-tuning stage. We then check again

Table 3.5: Summary and parameters numbers of the best DNN classifier model.

Layer (Type)	Output Shape	Param #
$Dense_1$	(None, 20)	360
$Dense_2$	(None, 25)	525
$Dense_3$	(None, 25)	650
$Dense_4$	(None, 25)	650
$Dense_5$	(None, 25)	650
$Dense_6$	(None, 25)	650
$Dense_7$	(None, 25)	650
$Dense_8$	(None, 25)	650
$Dense_9$	(None, 25)	650
Total parameters: 5,535		

the best model performance according to the Accuracy metric which allows us to evaluate the model results. Accordingly, the best prediction results are presented in table 3.6. This table shows the model's performance in relation to the classification rates of each target variable. According to the metrics defined by the equations (3.19a), (3.19b), (3.19c), and (3.19d), we can conclude that the quality of the classifier model is globally efficient and robust for automatic prediction of each state of the hydraulic components system. In particular, table 3.6 shows the Accuracy metric for the cooler condition is equal to 99.87%, valve conditions (99.60%), internal pump lake (99.09%), hydraulic accumulator (88.60%) and stable flag targets (94.17%). According to these results, the developed model is powerful in learning the failure states of the hydraulic system.

Table 3.6: Performance result of the proposed DNN multi-class classification model

TargetVariables	**Accuracy** (%)	**F1-Score** (%)	**Recall** (%)	**Precision** (%)
Cooler Condition	99.87	99.84	99.84	99.84
Valve Condition	99.60	90.40	99.16	91.69
Internal Pump Leakage	99.09	96.48	96.63	97.43
Hydraulic Accumulator	88.60	73.17	77.23	79.45
Stable Flag	94.17	90.91	90.00	92.44

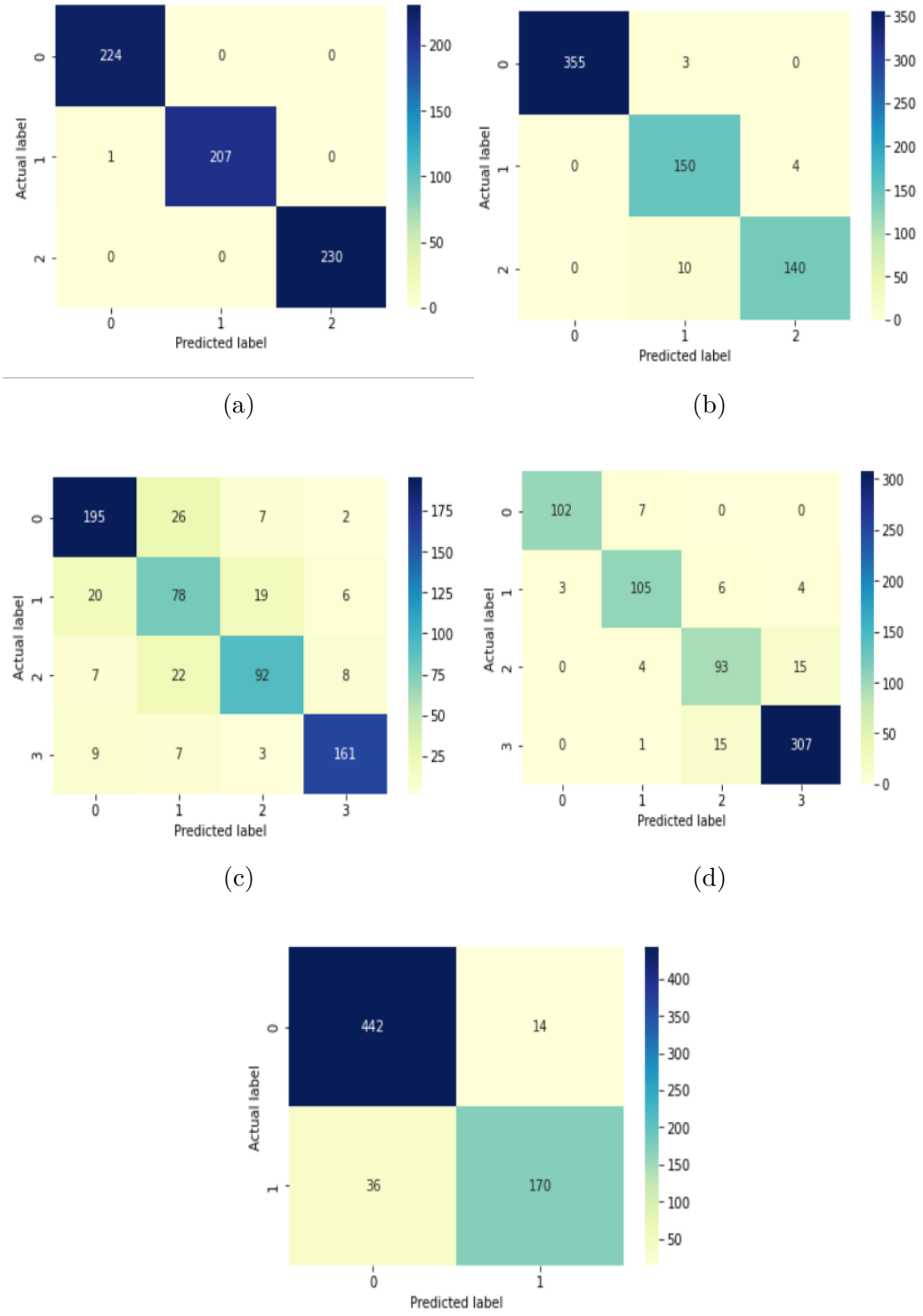


Figure 3.7: Confusion Matrices for the DNN Multi-classification module. Each matrix represents the frequency of the misclassification and their relative types of errors. Sub-figures (a), (b), (c), and (d) represent respectively the confusion matrix of the Cooler Condition, Internal Pump Leak, Hydraulic Accumulator, Valve Condition, and Stable Flag

An alternative method to evaluate the model's performance consists of analyzing the confusion matrix, which is the summary of the predicted model. These matrices (see Figure 3.7) show the frequency of misclassification errors of the models and their types for each hydraulic component condition. These types of classification errors compare the real classes with the predicted ones. Table 3.7 shows performance statistics, and by evaluating these tables, we can notice a difference between the accuracy of each subclass and the global classification accuracy of the target variable. In fact, when we focus on the condition of the hydraulic accumulator (Table 3.7(d)), we can deduce that the accuracy is not entirely sufficient. In particular, the slightly reduced pressure and severely reduced pressure state have respectively precision equal to 59% and 71%. Thus, it can be interesting to show that the degradation levels of these states are confused with each other. This refers to the states that are semantically linked among themselves, for example, the classes of slightly reduced pressure and severely reduced pressure are close. We have performed a performance comparison between the classification results of the DNN and the approach proposed by [196] which exploits a CNN. We can deduce that both models perform well and provide the same results globally. However, the significant difference between the two solutions is related to the explicability methods of the "black-box" models.

Table 3.7: DNN classifier model results for the multi-class classification of degradation levels of each state of the hydraulic system.

(a) Cooler Condition

	Precision	Recall	F1-Score
Close to total failure	1.00	1.00	1.00
Reduced efficiency	1.00	1.00	1.00
Full efficiency	1.00	1.00	1.00
Accuracy			1.00
Macro avg	1.00	1.00	1.00
Weighted avg	1.00	1.00	1.00

(b) Valve Condition

	Precision	Recall	F1-Score
Optimal switching behavior	0.97	0.94	0.95
Small lag	0.90	0.89	0.89
Severe lag	0.82	0.83	0.82
Close to total failure	0.92	0.92	0.92
Accuracy			0.92
Macro avg	0.91	0.90	0.90
Weighted avg	0.92	0.92	0.92

Table 3.8: DNN classifier model results for the multi-class classification of degradation levels of each state of the hydraulic system.

(c) Internal Pump Leakage

	Precision	Recall	F1-Score
No leakage	1.00	0.99	1.00
Weak leakage	0.92	0.97	0.95
Severe leakage	0.97	0.93	0.95
Accuracy			0.97
Macro avg	0.96	0.97	0.96
Weighted avg	0.98	0.97	0.97

(d) Hydraulic Accumulator

	Precision	Recall	F1-Score
Optimal pressure	0.84	0.85	0.85
Slightly reduced pressure	0.59	0.63	0.61
Severely reduced pressure	0.76	0.71	0.74
Close to total failure	0.91	0.89	0.90
Accuracy			0.79
Macro avg	0.78	0.77	0.77
Weighted avg	0.80	0.79	0.80

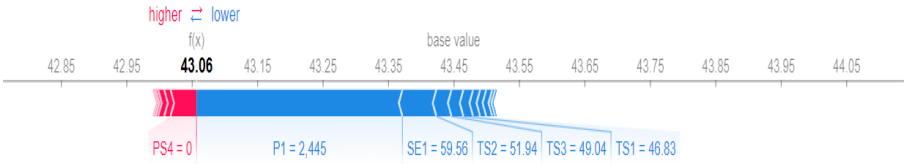
(e) State Flag

	Precision	Recall	F1-Score
Conditions were stable	0.92	0.97	0.95
Static conditions might not have been reached yet	0.92	0.83	0.87
Accuracy			0.92
Macro avg	0.92	0.90	0.91
Weighted avg	0.92	0.92	0.92

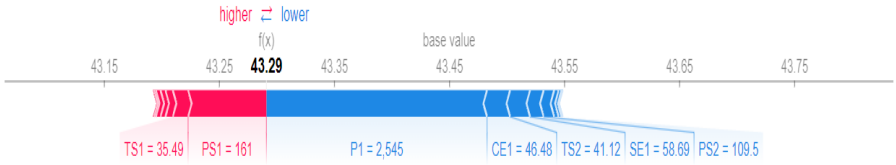
3.6.4 Deep SHapley Additive exPlanations Results

In this subsection, we focus on the main results related to the model's explanation. To highlight the importance, or the force of each x_i input variable, and to explain their role in the DNN decision-making, we applied the DeepSHAP approach to the developed model. As mentioned in the subsection (3.3.4.6), this approach attributes to each feature an importance value for a specific prediction. Furthermore, the Shapley value is indicated by an arrow that influences the prediction, thus, the positive Shapley value tends to increase the prediction, otherwise, the prediction decreases. However, these forces are expected to be balanced in the relevant prediction of the data instance. To illustrate the role of the features, we can observe

the Force Plot (Figure 3.8). In fact, sub-figures 3.8 (a) and (b) respectively show the local contribution of each i^{th} and j^{th} element at a given point of the prediction stages concerning the valve conditions. For the first Force Plot, we note that the base value is 43.08, and the predicted value is 43.06. Moreover, the feature PS4 contributes negatively and the features SE1, TS2, TS2, and TS1 positively contribute to predicting the variable state. Considering the second Force Plot, the base value remains the same, but the predicted value is 43.29. The features involved in the decision-making are subsequently P1, CE1, TS2, SE1, PS2 (positive contributions), and TS1, PS1 (negative contributions). Finally, a similar analysis will be conducted for additional target variables (Figures 3.9, 3.10, 3.11, 3.12) in section 3.7.



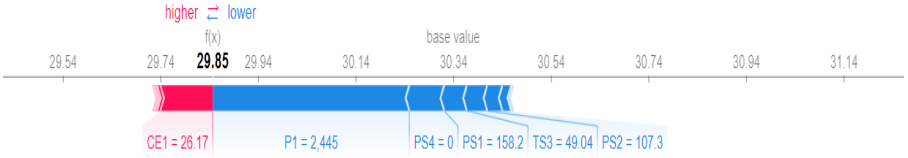
(a)



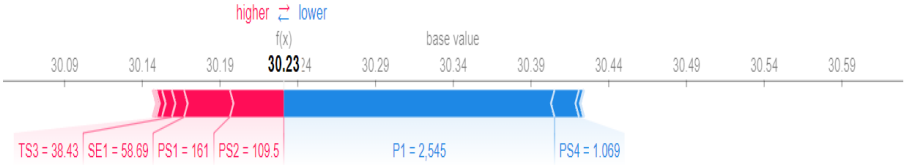
(b)

Figure 3.8: Force Plot and local interpretation for the valve conditions: This plot describes the function's output using the sum of these effects. Furthermore, the importance of the force of each feature is explained at different moments during the model training and decision-making. Thus, the features with a positive impact (contributing to the prediction being higher than the baseline value) are highlighted in red. In contrast, the features with a negative effect (contributing to the prediction being lower than the baseline value) are in blue. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN.

In addition to the local explanation, we can highlight that the purpose of the DeepSHAP approach is also to provide a global explanation of the features. It contributes to the choice of the model for predicting the states of each target variable. In this case, Figure 3.13 shows the global importance of features with arrows indicating their respective influence on the prediction. Thus, the summary Bar Plot (sub-figure 3.13 (a)) and the Summary Plot (Figure 3.13 (b)) respectively provide descriptions of the absolute average and the Shapley values of the force of each feature when predicting the target variable cooler condition. Furthermore, a positive Shapley value reflects a gain in the prediction, while a negative Shapley value indicates a loss in the prediction. Hence, they are revalued and updated in accordance with prediction requirements. Globally the most important feature contributing to the predicted operating conditions of the cooler state ('close to total failure', 'reduced efficiency', and 'full efficiency'), are the mean power motor (EPS1), cooling efficiency (CE1), temperature (TS3,

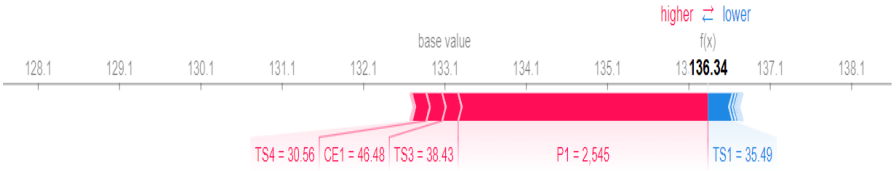


(a)

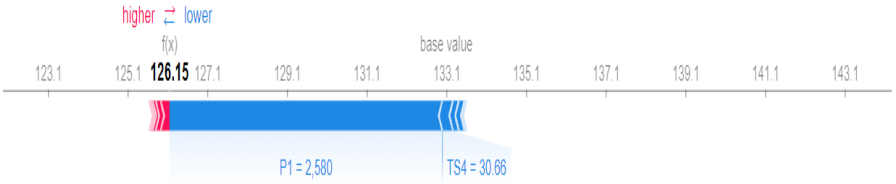


(b)

Figure 3.9: Force Plot describes the local interpretation) for the cooler condition classification state obtained from the DeepSHAP method module. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN.



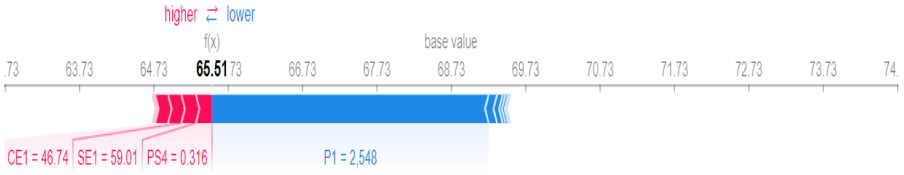
(a)



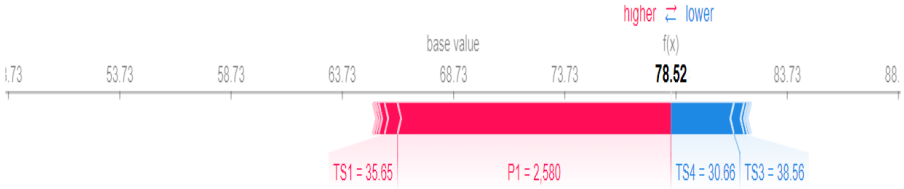
(b)

Figure 3.10: Force Plot shows the local interpretation for the internal pump leakage classification state obtained from the DeepSHAP method. In particular, sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN.

TS4), and pressure (PS1). As a result, the DNN multi-class classification model is trained only with these six features to predict the cooler state. The explainable module is, therefore, found to be an effective approach to performing the feature selection task.

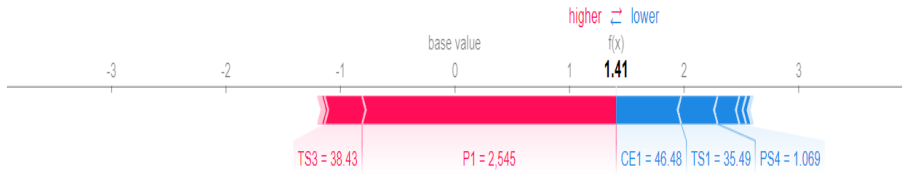


(a)

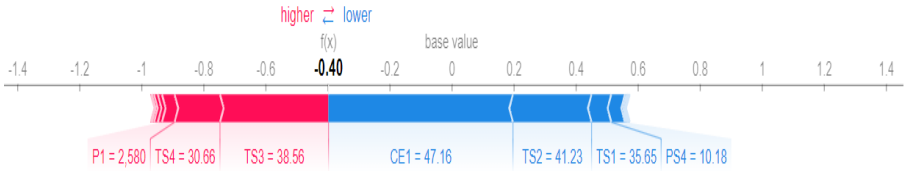


(b)

Figure 3.11: Force Plot is the local interpretation for the hydraulic accumulator classification state obtained from the DeepSHAP approach. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN.



(a)

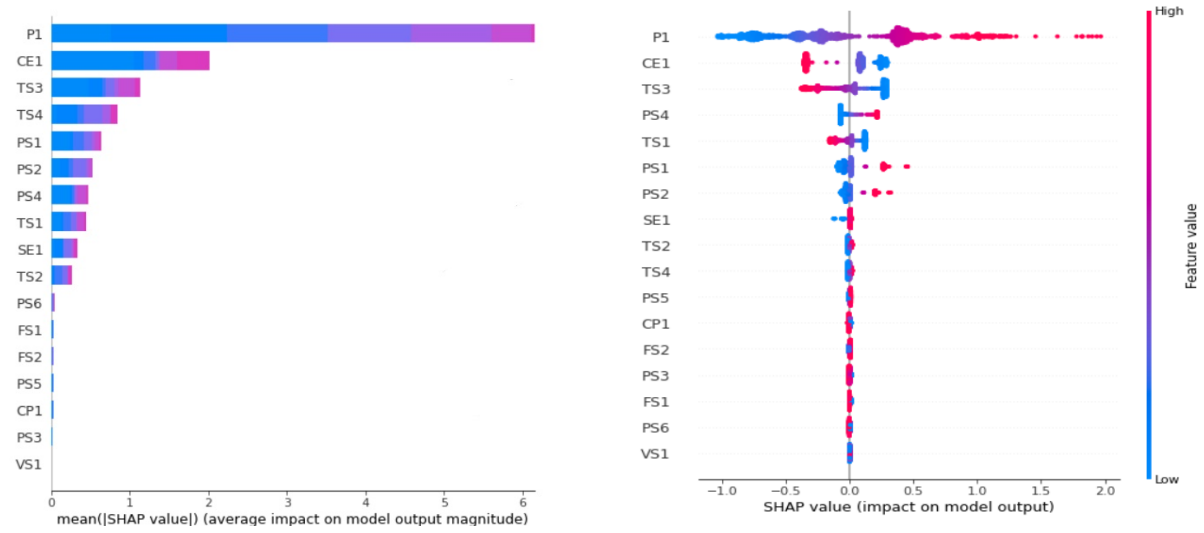


(b)

Figure 3.12: Force Plot represents the local interpretation for the stable flag classification state obtained from the DeepSHAP method. Sub-figures (a) and (b) represent respectively the local explanation of the i^{th} and j^{th} element in the decision-making process of the DNN.

3.7 Discussion

This chapter focuses on the condition monitoring of the hydraulic system. The aim is to create a framework that consists of developing a multi-class classification task combined with an Interpretable AI, the method applied to the hydraulic conditions. This can increase users' trust, or confidence in the performance, equity, or fairness of the "black-box" model. The input variables are described by the seventeen multi-sensor data and the target outputs are notably the cooler, internal pump leakage, valve, condition of the hydraulic accumulator, and



(a) (b)

Figure 3.13: Summary Plot (global interpretation) for the Cooler condition classification. The high Shapley values are in red dots, and the lower ones are in blue. The figures show the global importance of the variables. In Particular, sub-figure (a) illustrates the absolute average of the Shapley values, and sub-figure (b) shows the Shapley values of each feature according to their order of importance.

stable flag. To respond to this purpose, we have proposed a detailed framework composed of two modules: the fully connected DNN classifier model and the DeepSHAP explainable approach. Data pre-processing operations of formatting, Min-Max normalization, and One-Hot-Encoding applied to the target variables have been considered. Since some of the target variables have slightly unbalanced classes, we use the re-sampling and Cross-Validation techniques. Regarding the descriptive analysis, there are no existing missing data in the considered data sets. Furthermore, the histogram (Figure 3.6) shows that the input features do not fit with any distribution, thus, we cannot make any assumptions about the possible presence of outliers on the data sets. The performance of the classification model is evaluated according to the confusion matrix (Figure 3.7) and metrics such as Accuracy, F1-Score, Recall, and Precision (tables 3.6 and 3.7). The main DeepSHAP results are presented by the local contributions or Force Plot (figures 3.8, 3.9, 3.11, 3.10 and 3.12) and the Summary Plot or global contributions (see figures 3.13, and 3.15).

Regarding the classification result, table (3.6) presents the qualitative classification results, notably the accuracy for each classification of target variables. The cooling condition (99.87%), valve conditions (99.60%), pump leak (99.09%), and stable flag (94.17%) have a classification performance rate near 100%. However, the result of the hydraulic accumulator (88.60%) is less accurate and more difficult to obtain. To improve the understanding of these results, we analyze the misclassification rate and the explainer model for each target variable with their respective sub-classes. The first hydraulic state is the cooling condition, which is an important issue for the performance of the hydraulic system [174]. The results obtained show

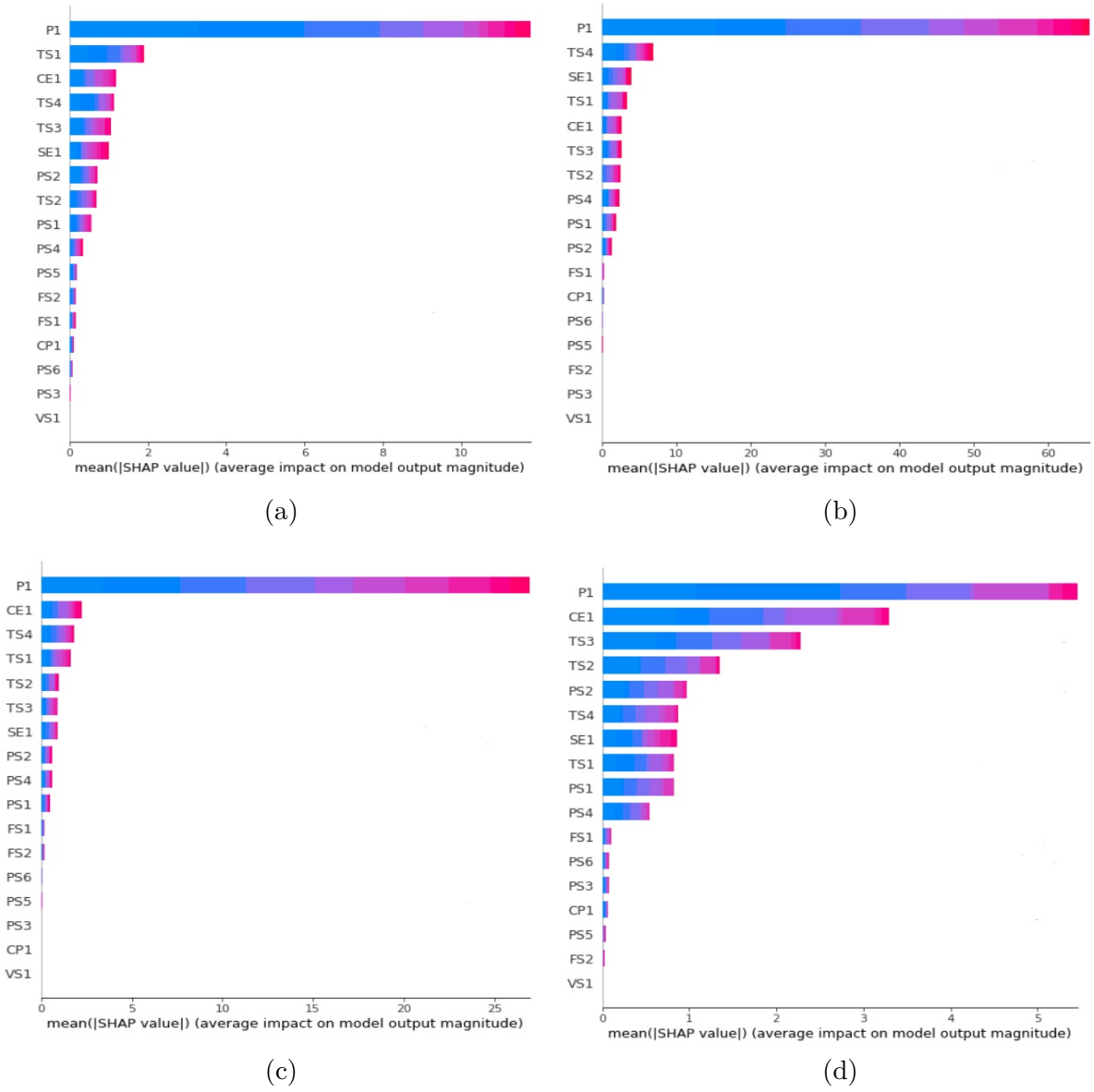


Figure 3.14: Summary Bar Plot represents the global importance of each component of the hydraulic conditions. In particular, sub-figures (a), (b), (c), and (d) show respectively the global importance feature or contribution of each sensor in the decision-making models for the components of the hydraulic system (Valve conditions, Hydraulic accumulator, Internal pump leak and Stable flag)

that the model has an accuracy rate equal to 99.87% and each subclass is perfectly classified. Despite these satisfactory results, we exploit DeepShap to explain the impact of each input variable in the decision rule of the model. In this regard, we undertake an in-depth analysis of the local (Figure 3.9), and the global contributions (Figure 3.13) for each feature. Considering the local explanation, we can observe that each sensor contributes differently to the learning

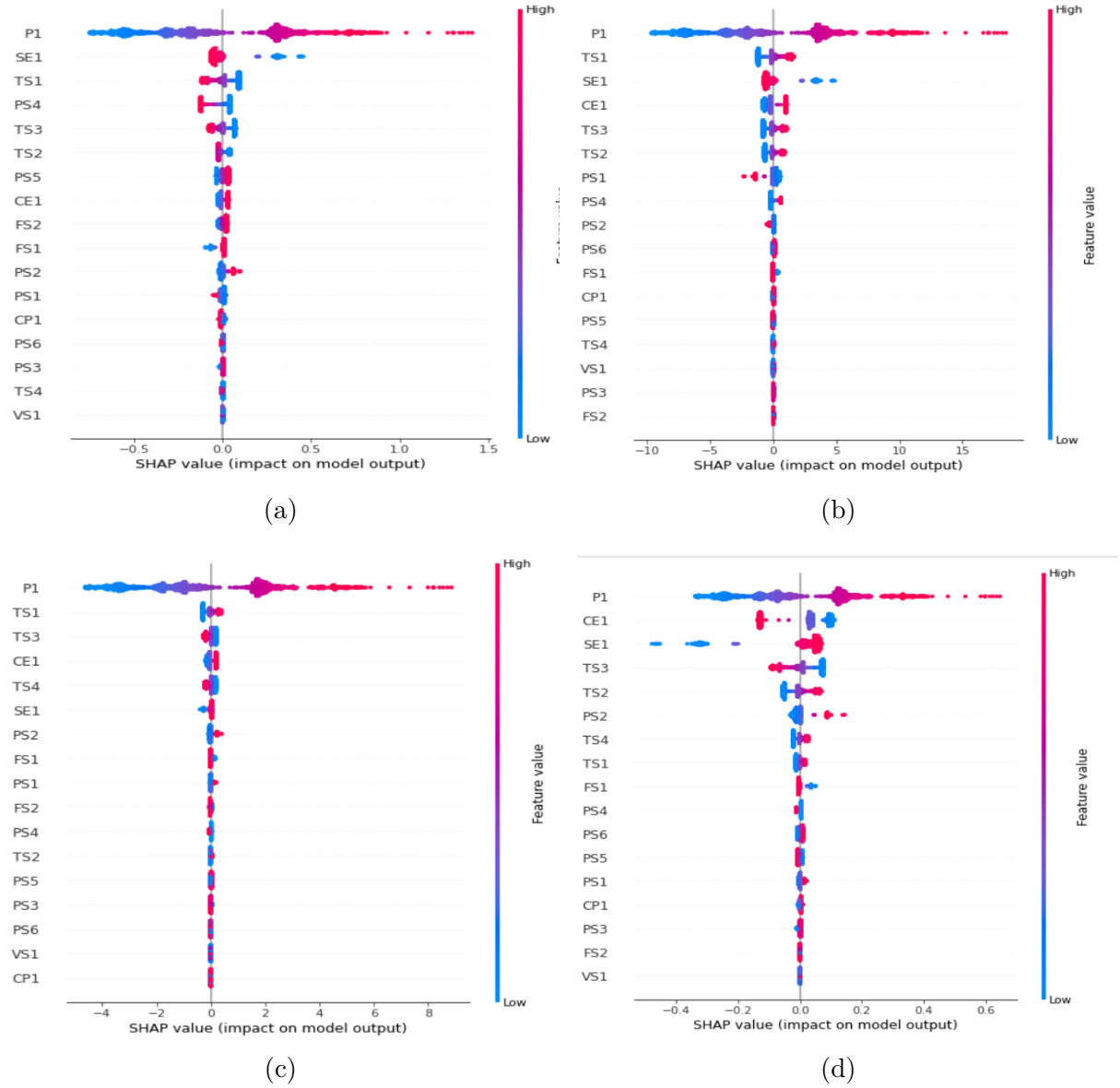


Figure 3.15: Summary plot: Sub-figures (a), (b), (c) and (d) show respectively the global importance or contribution for each sensor/feature contribution to the decision-making DNN model.

process according to their weights or Shapley values ($f(x) = 29.85$, $f(x) = 30.23$). However, the baseline value (30.34) does not change. Furthermore, if we consider a given moment of the model's learning process (sub-figure 3.9 (b)), we can observe that the cooling efficiency sensor (CE1) has a negative contribution while the cooling power P1 (mean value of the motor power EPS1), pressures (PS4, PS1, PS3), temperature (TS3) of sensors have a positive contribution. However, these contributions could be different compared to the previous moment of the model's learning process (sub-figure 3.9 (b)). Considering the local explanation, we

can observe that each sensor differently contributes to the learning process according to their weights or Shapley values ($f(x) = 29.85$, $f(x) = 30.23$), however, the baseline value (30.34) does not change.

Regarding the global contribution (Figure 3.13), it is noticeable that the sensors P1, CE1, TS3, TS4, PS1, PS2, and PS4 have a major impact on the model decision-making. These findings demonstrate that the hydraulic system cooling condition is most probably conditioned by the quantity of cooling pumped, pressure, power of the motor, and temperature of the cooler to maintain the pump at a normal temperature. The second hydraulic state is the valve condition. As for the previous case, the accuracy rate of the classification model is near to 100%. In addition, the subclass precision rates such as 'Optimal switching behavior', 'Small lag, severe lag' and 'Total failure' are respectively 97.87%, 90.87%, 82.87%, and 92.87%. Sub-figures 3.8 (a) and (b) represent the local explanation and we observe that the first force plot illustrates that the sensor (PS4) has a positive force, whilst sensors (P1, SE1, TS1, TS2, TS3) have a negative force. On the other hand, when we focus on the second Force Plot, the following couple of sensors (TS1, PS1) have a negative force, whilst (P1, CE1, TS2, SE1, PS2) have a positive force on the decision-making of the algorithms. In addition, based on Summary Plot (sub-figure 3.14 (a)), we show the most important sensors (P1, TS1, CE1, TS4, TS3, SE1, TS2) have global importance in the model training. Finally, we can also conclude that the operating states of the valve conditions are strongly conditioned by the volume flow, power motor, temperature, and pressure measured by the following sensors (P1, TS1, CE1, TS4, TS3, SE1, and TS2). Moreover, the valve commutation performance can affect the flow of the hydraulic fluid.

The internal pump leakage is the third hydraulic state, the results of the model give an accuracy rate equal to 99.09%. However, when we observe the misclassification result of each subclass ('No leakage', 'Weak leakage', and 'Severe leakage') we find 100%, 92%, and 97% respectively, so, this demonstrates that the results globally remain significant. We are also interested in analyzing the local explanation. Figure 3.10 (a) represents the Force plot where the sensors (TS4, CE1, TS3, P1), and TS1, therefore, have a negative and a positive force at the given learning moment. However, the contribution changes when we focus on sub-figure 3.10 (b), so the sensors P1 and TS4 have an opposite contribution. In terms of the global contribution, Summary Plot (figures 3.14 (c) and 3.15 (c)) shows the list of sensors (P1, CE1, TS4, TS1, TS2, TS3, and SE1) classified in order of their importance. As in the case of the valve condition, the volume flow rate is important, because if there is a leak in the pump, this anomaly can have a negative effect on the flow rate of the hydraulic fluid. In addition, the efficiency factor is important, thus, a large leak could affect the general efficiency of the system.

Classifying the condition of the hydraulic accumulator conditions is more challenging, as, its accuracy rate (88.60%) is less than other hydraulic system states. Nevertheless, by observing table 3.7(d), the misclassification of the 'Optimal pressure' and the 'Close to total failure states' are respectively 84% and 91%. If we concentrate on the precision rate of the 'Slightly reduced pressure' (59%) and the 'Severely reduced pressure' states (71%), we can note that their precision is not satisfactory. Consequently, it is interesting to show that

during the learning process, the model tends to confuse these two sub-classes, since they are semantically close. The reasons for this weak performance may be related to the quality of the data.

In general, the data collection process represents a major industrial challenge. This can be explained by the fact that the sensors or machinery may not generate qualitative or quantitative data to identify the real state of the system deterioration. This problem may also be related to the accuracy of sensor measurements between consecutive degradation states. To address this issue, the manual labeling of raw data can be performed by an expert. However, this process can be time-consuming with additional operational, and economic costs. It may also be subjected to mislabeling risks. Alternative strategies have been proposed in [151] such as performing artificial re-sampling, data augmentation, clustering, and simulation of trusted or fake data using Generative Adversarial Network (GA) and Transfer Learning approaches.

Concerning the DeepSHAP results, figure 3.11 and sub-figure 3.14 (b) respectively represent the local and global contribution of each sensor to the prediction of the hydraulic accumulator's condition. In addition, sub-figure 3.11 (a), shows the local explanation or the Force Plot for the prediction of this target variable obtained. Thus, the base value is 65.51, and the value of the function $f(x)$ is equal to 68.73. We can notice that the sensors (CE1, SE1, and PS4), have a positive contribution and that P1 has a negative contribution to the model classification. However, if we capture a further moment during the decision-making of the algorithm (Figure 3.11 (b)), we observe that the contribution is not the same as in the preceding case. So, the sensors (TS1, P1), and (TS4, TS3) respectively have a positive and a negative contribution. However, for the global contribution (see sub-figures 3.14 (b), and 3.15 (b)), we note that the most important features are P1, TS4, SE1, TS1, CE1, TS3, and TS2. We have mentioned that the prediction of this target variable is more challenging. It is partly related to the distribution of their sub-classes that are often semantically confused with each other. To solve this issue, we have challenged several approaches to optimize the DNN classification model. The first method consists of making a model more complex by increasing the number of hidden layers or epochs. However, this approach had no significant results, and the accuracy rate remained the same as before. The second approach was to exploit the information provided by the importance of the global feature of the explainer DeepSHAP module. This consists of using the most important sensors to select features (PS4, FS1, FS2, TS1, TS2, and TS3) for retraining the DNN classifier model. Thus, we achieve excellent results, the accuracy rate (66%) of the 'Slightly reduced pressure' state increases and it is an improvement on the previous case. Although this result remains relatively low, there has been a real improvement in the use of the new approach. Finally, the results obtained are still considered satisfactory when they are compared to those provided by the previous paper [174].

The last hydraulic system state is the Stable flag condition. In this case, we have the unique target variable for which the classification is binary. The most important result shows that the accuracy rate is equal to 94.17%. Globally the precision of the 'Conditions were stable' and the 'Static conditions might not have been reached yet state' is equal to 92%. In addition, the global explanation (see sub-figure 3.15 (d)) indicates that the most relevant

sensors on decision-making models are notably P1, CE1, TS3, TS2, PS3, TS4, T1, and PS1.

3.8 Conclusions and Future Work

This chapter presented a detailed framework for condition monitoring based on hydraulic systems and multi-sensor data. The main finding addressed is the prediction of the hydraulic conditions, and the explanation of the model developed. To investigate this issue, we have developed two main modules: A multi-class DNN classification model combined with a DeepSHAP XAI approach. The resulting model of the first module is efficient and robust for the classification of the hydraulic system conditions such as the cooler, valve, internal pump leakage, condition of the hydraulic accumulator, and stable flag. In addition, the second module named DeepSHAP provides further information about the local and global importance or contribution of each feature to the model's decision-making. The main contribution of our proposed approach compared to other studies is the capacity of the DNN classifier model to operate directly on the data without performing the features selection techniques, while still capturing the deep aspects of the features. Furthermore, in comparison to other XAI techniques, the DeepSHAP discriminates more effectively in the network model outputs and provides both local and global explanations.

The DNN classifier model has been evaluated using several metrics such as Accuracy, F1-Score, Precision, Recall, confusion matrices, and misclassification which are more robust in dealing with prediction problems with the unbalanced classes. Using the above metrics, the results demonstrate that the classification rate of each target variable is efficient, in particular cooler conditions (99.87%), valve conditions (99.60%), internal pump leakage (99.09%), and stage flag (94.17%), with except for the hydraulic accumulator conditions (88.60%). To understand the cases for which the model is not accurate, we have exploited the confusion matrices. We observe that the algorithm is less efficient when the deterioration state levels labeling are too fine or are semantically similar. To address this issue of data quality, several solutions are available including manual data labeling, artificial re-sampling, data augmentation, clustering technique, fake reliable data simulation, or transfer learning.

We have also focused on the explanation of deep learning reasoning using the DeepSHAP approach. The main result of the explanation approach shows the local and global contribution or importance of each sensor in the decision-making process of the DNN model. Furthermore, this approach provides a better explanation of the classes of the model output. We use these results to optimize the prediction model. The DeepSHAP module is more in line with human intuition since it provides means for improving the understanding and interpretation of practitioners about predictive reasoning. Initially, we built the DNN model on the complete set of sensors, and the DeepSHAP approach assisted us in the selection of the most relevant sensors to retrain the model. We found that by reducing the number of features from 17 to 6 the accuracy of the hydraulic state increases. This suggests that not all sensors are necessary to understand hydraulic accumulator conditions. In addition to the high-performance classification, this framework helps to support all stakeholders in their un-

derstanding of the decision-making process. This can promote trust or increase confidence in the use of Condition Monitoring applications based on Artificial Intelligent models.

Limitations and Future Work Orientation

A limitation of the developed approach is the fact that the DNN model does not consider sensor failures or incomplete data. For future studies, it will be interesting to investigate the robustness of DNN models for missing or noising data. Moreover, we must ensure that our data management and predictive models have the opportunity to consider abnormal sensor behaviors such as aging or mislabeled data. In addition, we will perform a classification model using more data, such as air, oil, and water contamination data, which are one of the main causes or factors of hydraulic system failure. Regarding the explanation of the model, the combination of several eXplainable Artificial Intelligence approaches likely to provide the best results should be considered. Moreover, it would be necessary to confirm the results of the framework with domain experts.

In addition to the explanation framework, there is additional interest in the generalization of traditional methods. After providing more detailed information on explanatory approaches, we will develop Physics-Informed Neural Networks. This chapter illustrates how to extract knowledge (decision-making rules) via an explainable model. However, in the following chapter, we will introduce knowledge or constraints during neural network training. This knowledge helps to guide or inform the NN to follow the process topology. The PINN model is designed to be adaptable to a variety of processes and to capture local and global information. The resulting hybrid framework is generalizable, robust, and accurate in predicting optimal model parameters and output data.

Chapter 4

Physical-Informed Neural Networks and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding (FSW)

Sommaire

4.1 Introduction . . . 103
4.2 Overview of the Process Modeling Techniques . . . 106
4.2.1 Numerical Simulations Approaches . . . 106
4.2.2 AI-based models . . . 107
4.3 Thermomechanical Process: Friction Stir Welding (FSW) . . . 110
4.3.1 Principle and Operating Mode . . . 110
4.3.2 Parameters and Main Defects of the Process . . . 111
4.4 Theories of the Numerical Simulation of the Thermomechanical Problem . . . 111
4.4.1 Conditions of the model . . . 111
4.4.2 Geometrical modeling . . . 111
4.4.3 Meshing and boundary conditions . . . 112
4.4.4 Governing equations . . . 113
4.4.5 Parameters and Mechanical properties . . . 115
4.5 Proposed Methodology . . . 115
4.5.1 Data Pre-Processing . . . 116
4.5.2 Type of Layer Activation Functions of the Feed-forward Neutral Networks 116
4.5.3 Regularization and Physics-based Loss function . . . 119
4.5.4 Observations about the Combined Loss Function . . . 124
4.6 Results and Discussions . . . 127
4.6.1 Result of the Numerical Simulation of the 2D Data . . . 127
4.6.2 Physics-Informed Neural Networks Trained on the FVM Solution . . . 127
4.6.3 Discussions . . . 132
4.7 Conclusion . . . 133

abstract

In recent years, Artificial Intelligence (AI) techniques have seen a significant rise in popularity through their performance. These techniques are used in many industrial applications such as modeling, identification, optimization, prediction, and control of complex systems. AI-based models are also developed in wide applications related to thermomechanical Friction Stir Welding (FSW). This chapter focuses on a new class of Neural Networks (NNs), that combines automatic Learning and physical laws known as Physics Informed Neural Networks (PINN). A numerical simulation of the FSW process has been developed using the Finite Volume Method (FVM) and the industry-leading computational fluid dynamics (CFD) simulation software called Ansys Fluent". The simulation results are used as data for training and validating the hybrid model that we have developed using the Pytorch library. The FSW process is a highly computationally time-consuming process. In addition, it can be difficult or impossible to simulate the whole physical duration of the process or to reach the stationary regime. Moreover, in the literature PINNS models are often used to address problems in fluid mechanics and to resolve nonlinear Partial Differential Equations (PDEs). We develop a Fully Connected Neural Network (FCNN) informed by the physical law that addresses a solid mechanics problem where the viscosity expressed by the modified NSE represents the strain rate of the material which depends on both the loading intensity and the loading rate. However, this NSE is more complex and requires higher-order differentiation. The obtained PINN is a supervised learning task while satisfying a given physical law described by NSE. In addition, the model aims to learn the transient phase of the process. By including the knowledge of the physical constraint or the regularization terms, we obtain a composed loss function. This knowledge aims to understand the process and forces the model to follow the defined conditions imposed by the physical constraint, so the obtained model is generalizable. The major results of this framework have shown that once trained the PINN model can be a valid substitute for the numerical models, thus allowing big time-saving thanks to their memory effect and making it possible to find an approximate solution to the PDEs which is impossible to solve analytically. Using the Root Mean Square Error (RMSE), we conclude that the proposed framework is more robust and has a high ability to predict the whole process duration (transient and stationary) than the traditional Artificial Intelligent (AI) approaches.

Keywords: *Physics constraints, Physical Informed Neural Networks (PINN), Loss function, Hybrid modeling, Friction Stir Welding (FSW), Numerical simulation, Regularization Approach, Fluid mechanics, Partial Differential Equations (PDEs), Navier-Stokes equation.*

4.1 Introduction

In recent decades the mechanical engineering industry has increased the use of composite materials in the field of transport or aircraft manufacturing [238]. However, the difficulty of assembling these materials by using conventional liquid-phase welding techniques remains a crucial and challenging problem. To overcome these issues, the Friction Stir Welding (FSW) technique has been introduced in 1991 by the Welding Institute in England [239]. This solid phase process offers several advantages in addition to the quality and robustness of the joints, the process is energy-efficient, environmentally friendly, and versatile. This process involves complex local thermomechanical phenomena in the vicinity of the welding tool which are not visible to monitoring cameras (high-speed cameras, infrared cameras). Handling these local phenomena allows for avoiding FSW defects like the flash, surface seizure, vacuum bonding, kissing bonds, and onion rings [240]–[242]. The experimental study of this process requires heavy investments or expensive services. To anticipate these anomalies several studies have shown that process monitoring is prominent for better optimization of the production line [243], [244]. Optimization is important in the engineering industry in reducing production costs and time consumption. Thus, the significant void defects can be implicitly detected by observing monitored forces, however, most of the monitoring systems focus on the global variables (i.e., machines' power consumption, forces in more directions). Furthermore, the quality of the welded parts is related to local phenomena which are complex to be monitored and detected. In addition to this, the transition phase where various parameters are involved is difficult to model. In addition, the experimental study of this process requires heavy investments or expensive services. To understand the process such as the role of parameters, the impact of the high variations of the velocities, pressure to the torque or the forces, the welding conditions, and the temperature distribution we can use the numerical simulation approaches [245], and Artificial Intelligence (AI) techniques [246]. The quality of FSW joints is greatly influenced by a number of parameters. In this study, we focus on the impacts of Rotation velocity (rev/ min), Velocity of advance (mm/min), Number of mesh nodes, and Time step (millisecond) and the variables to be predicted are respectively the velocities and pressure.

The numerical simulation technique has become an important and popular area of research [239], [247]–[250]. It helps to better understand the physical phenomena related to the complex process. It allows the estimation of different process parameters (geometry, tool speeds, etc.). The numerical modeling techniques have been exploited to explain and predict important features of the physics of the processes involved in the FSW process [245]. The technique presents significant results, in particular, it has been used to model the different profiles and spindle speeds [251]. In addition, it was deployed for modeling heat transfer [252], for modeling metal flows [253], and for coupled modeling between viscoplastic flow and heat transfer (predicting temperature and residual stress distributions), [254]. The numerical solution can be time-consuming and requires a high GPU (Graphics Processing Unit) or CPU (Central Processing Unit) memory [239], [247], [248]. In the process, a strong thermomechanical gradient at high speed operates, which requires fine mesh elements and small time increments [255]. This constraint facilitates a better simulation and leads to an increased dramatically

memory cost. Thus, the simulation is repetitive and straightforward involving the repeated approximation of large systems. To improve computational efficiency, several approaches have been proposed. The authors of [249], [256] have used the transient simulation to limit the number of experimental case studies and computational time. However, simulation is not an easy task since the FSW process is a very complex process. It involves the interaction of several strongly coupled and non-linear thermal and mechanical phenomena (i.e., plastic deformation, material flow, heat generation, surface interaction between tool and workpiece). This process is difficult to simulate due to the high deformations generated by the different parameters during the mixing stage of the transient regime (interpolation of the velocity fields and complexity of the tool parameterization). In addition to this, the technique requires the use of advanced multi-physics solvers to estimate the parameter. Furthermore, as real or quasi-real-time simulation data are required for process monitoring, numerical simulations cannot be used directly.

In practice, the estimation or identification of the optimal parameters of the process requires a large number of simulations, which is costly. An alternative approach is to develop AI-based models. In recent years, AI has dramatically changed the manufacturing and materials industries [151]. It is used in many industrial applications like image classification, handwriting recognition, speech recognition and translation, and computer vision. It can be applied to address issues of the optimization, quality control, and prediction of failure modes [257]. Figure 4.1 shows the different applications of AI.

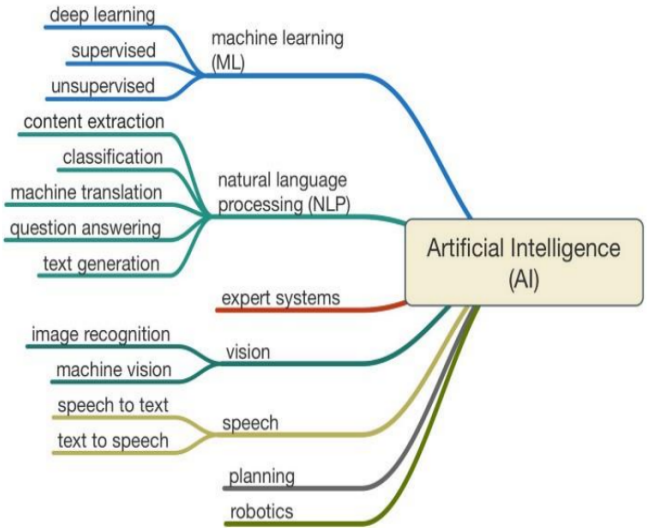


Figure 4.1: Illustration of the number of possible ways in which AI can be performed [246].

Paper [258] highlights the applications of AI models for predicting FSW weld states, while SVMs are used to predict the maximum temperature of a weld [259]. Authors [260] combine SVMs and ANNs to classify and locate defects in the weld. Papers [260] and [258] present respectively a convolution neural network (CNN) for defect detection and an artificial neural network (ANN) for monitoring fracture failure. Paper [261] presents a Theory-Guided Data

Science model for approximating the PDEs using a parametric function or a regularized loss function. When the model is sufficiently trained, the approximate solution of the PDE can be obtained whatever the input variables are used in the analysis. However, their accuracy is highly dependent on the ability to simulate a large and highly reliable amount of training data that represents the phenomenology of the physical system. We observe that the proposed solutions do not address the issues of the numerical simulation models which require significant computational time and an important memory to estimate the model parameters. In addition to this, the calibration of these parameters can affect the performance and quality of the model. Despite the numerous uses of AI techniques in the industry, ANN can often generate complex models that do not fully explain the investigated physical phenomenons with a significant generalization error.

Despite the performance of ANNs, hybrid models trained with ANNs tend to perform better and more accurately. In this perspective, we introduce a new paradigm known as the Physics-Informed Neural Network (PINN). This framework has been developed to train the FCNN with the known equations that govern the physics of a system. The training of the FCNN is performed with a cost function penalized by constraints, and initial and boundary conditions. A traditional application of PINNs is to solve systems of Ordinary Differential Equations (ODEs) and partial differential equations (PDEs) by estimating their various parameters. In this case. The approach is used to provide a better insight into the FWS process, and to estimate the optimal parameters. In the next section, we provide more details about AI-based approaches, focusing on PINN models and their applications.

Objectives of the Study

In recent literature, PINNs models are often applied to fluid mechanics problems. However, the proposed PINN model addresses a solid mechanics problem where the viscosity is expressed by the NSE. This equation is much more complex and requires higher-order differentiation. Knowing that FSW is a computationally time-consuming process, it is difficult, or even impossible to simulate the whole physical duration of the process or to reach the stationary regime. The model developed aims to learn the process including their transient regime. In addition, it should be able to predict the whole process duration like transient and stationary regimes. We train and validate the framework on the synthetic data generated by the FVM method. The dataset contains the input variables are the spacial coordinate (x, y) and time t while the output variable is velocities (v_x, v_y) and total pressure p. These data satisfy the governing equations with a different level of accuracy, and the error can be viewed as white noise in the data. Moreover, when the data are preprocessed correctly, the PINN training converges quickly to the optimal solution and parameters. This result is valid for the case where the data are generated by the FVM method with coarse mesh. This result shows the robustness and generalization of this approach. Furthermore, by adding physical constraints in the loss function, the training model converges on very sparse data. Lastly, the developed framework must be able to substitute the numerical simulation models of thermo-mechanical processes. This aims to provide rapid predictions and investigate various process

parameters. The rest of this chapter is organized as follows: section 4.2 addresses the overview of the modeling approaches including the numerical simulation models and AI-based models. The description of the FSW process is presented in section 4.3. In section 4.4, we focus on the theory of the numerical simulation of the thermomechanical problem. Furthermore, section 4.5 presents the research methodology and the key contributions. We explain the most important results and discussion in section 4.6. We make conclusions in section 4.7

4.2 Overview of the Process Modeling Techniques

4.2.1 Numerical Simulations Approaches

Although FSW welding has emerged as the state-of-the-art joining process, the welds require special inspection [243], [250]. The microstructure of the joints reveals elements that are indicators of weld quality. The FSW numerical simulation is a popular area of research because the physics of the process is complex and requires the use of advanced multi-physics solvers. This method has several advantages, such as its simplicity, and large application areas. There are several approaches that help the numerical analysis models for FSW, including, the Finite Difference Method (FDM) [250], the Finite Element Method (FEM), and Lagrangian-based FEM [262] are used to develop in the numerical simulation model for the heat transfer, temperature, and distribution. In particular, the FEM method consists in determining local fields to be assigned to each element so that the global field obtained by juxtaposing these local fields is close to the solution to the initial problem, however, these methods suffer from major distortions. The authors of [263] exploit the finite volume method (FVM) with the Eulerian approach to analyze and simulate the flow of matter, but it does not make it possible to follow the evolution of each material point. The Arbitrary Lagrangian-Eulerian (ALE) is a mesh simulation method that includes material advection from the Lagrangian mesh into an Eulerian mesh. This approach helps to simulate the higher levels of plastic deformations. However, ALE suffers from certain drawbacks, thus, the simulation of the FSW process is time-consuming and is subject to precision errors. There are also measles simulation methods such as Smoothed Particle Hydrodynamics (SPH) [264] that help to follow the whole FSW process with very low plastic deformation and mesh distortion, however, it is expensive in terms of computing time. The authors of [265] have shown that this time can be improved with the use of parallelization and graphics processing unit (GPU) methods to considerably improve computing time. In addition to this the authors of [255], have shown that to understand the process and obtain correct results, it is essential to use a numerical simulation by the fine mesh method. They assume that the size of the elements is less than or equal to one-tenth of the feed per revolution of the tool lobes. This last requirement explodes the calculation times and makes this process not simulated by means available today. To overcome the computational time problems, numerous studies have developed AI-based models [250], [266], [267].

4.2.2 AI-based models

In recent years, AI approaches have become very popular in light of their technological progress and numerous applications including complex mechanical systems [76], [268], [269]. Thus, in the literature, there are several frameworks or approaches for modeling these systems [151]. Figure 4.2 presents these approaches. Physics-based modeling focuses on the exploitation of physical constraints or knowledge. This knowledge represents the main source of information while the use of data is marginal. The data-driven modeling highlights the use of data (the main source of information is the data) and the axis regarding the exploitation of physical constraints or knowledge is insignificant). Hybrid modeling uses both data and physical knowledge. Thus, it extracts information from both information sources. In the rest of this sub-section, we provide detailed descriptions of these approaches.

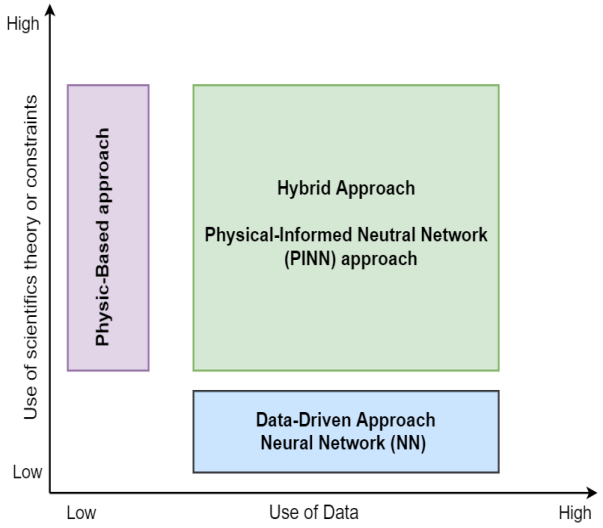


Figure 4.2: Schematic representation of the approaches used to address predictive tasks on a complex system.

(a) The first model is known as physics-based modeling, this approach requires the construction of a dynamic model by integrating various constraints or degradation linked to the physical processes [67], [68], [270]. This model can be divided into two groups; (i) Physics-based equations [271] that represent the relationships between different physical variables. This method helps to validate the data simulation process and to understand the effect of various physical components [272]. For example, these components are profiles, the tool spindle shapes on temperature, the stirring zone, and the power consumed by the welding. (ii) The numerical simulation model (see subsection 4.2.1) is presented in the form of PDEs which is based on the physical laws (i.e., Burgers', NSE, Wave, Laplace's, or Poisson's equation). The resolution of these equations is obtained via numerical simulation methods [250], [262]. The main advantages of the physics-based approach are related to the model parameters, the degradation or the deformation of the process can be explained and associated with the

variations of these parameters. Moreover, the results can be easily interpreted. However, physics-based modeling is limited to its ability to extract knowledge or insights directly from data that is mostly based on available physics. Furthermore, the models generated are often too complex leading to incorrect results. In the case of similar processes, the solutions obtained are very often proved to be not generalizable, or transposable [69], [70]. In addition to this, the approach requires a high computational time and significant memory GPU. Furthermore, for parameter estimation, the calibration of the model parameters is a difficult task due to the combinatorial nature of the search space.

b) The second approach is known as Data-Driven modeling. This framework uses heterogeneous data to extract some relevant features, knowledge, or patterns useful for characterizing the system. It is applied in many applications [273]–[276] such as anomalies detection, computer vision, image processing or natural language processing [201], [202]. It can be classified into several approaches: Machine Learning (ML), Statistical, Stochastic, or Deep Learning (DL) models. We can exploit the possibility to develop an FCNN with many hidden layers, this is efficient for non-linear problems and complex relationships. In addition, the model helps to extract some relevant features that facilitate the process compression, due to articulated architectures of different transformations. The authors of [268], [277] propose a regression model to explain the direct relationship between the temperature, velocities state of the tool, and its forces. However, this model does not provide a good performance due to the relationship between these variables is always non-linear. In addition to this, the approach suffers from several issues, such as the model instability based on incomplete or noisy data [278], large-scale learning problems [279], the curse of dimensionality, and the overfitting phenomenon [280]. To address these issues, the researchers have developed some powerful frameworks [281]–[284] such as Convolutional Neural Networks (CNN), Restricted Boltzmann Machine (RBM), Artificial Neural Networks (ANN) Auto Encoders network (AE) and Generative Adversarial Networks (GANs). The authors of [267], [285], [286] have developed an ANN to predict the FSW parameters of aluminum plates and their mechanical properties. Papers [287] present a conditional GAN (cGAN) for the prediction of the optimal temperature distribution.

Unlike physics-based modeling, this approach is faster and can be used for real-time prediction. Despite their performances, the models are limited to global observation and cannot predict small local void defects in the weld seam. In some situations, this approach model loses interest mainly due to its instability and capacity to capture new changes related to the process. This can be caused by the collected data that does not automatically represent all the process states or in presence of noisy data [288]. The estimation of the network parameters can be difficult because the learning complexity lies in the disappearance of the gradients with the increase of the number of hidden layers. An optimal architecture and a number of parameters must be determined to avoid the problem of overfitting. In addition to this, when the data are biased or low-dimensional, the efficiency of these models can be limited, as they can lead to not sufficiently robust and non-generalizable results. (c) The last approach is a new paradigm known as hybrid modeling. This approach is a combination of knowledge-based or physics-based or data-driven modeling [289]. Author [290] use a hybrid model called Cellular Automata Finite Element (CAFE) combined with Artificial Neural Network

(ANN), to predict the evolution of grain size and yield strength during FSW. To address the limitations of the previous approaches we introduce any form of prior knowledge about the physics of the problem in the learning algorithm. This knowledge is also included in the data simulation, to provide a kind of "training guide". This is performed by using an expanded loss function; thus, the network output is constrained to satisfy a system of PDEs using a regularization function. In this regard, the algorithm imposes a penalty to force the process solutions to converge as fast as possible to the correct solution. In this case, the model adapts continuously to the operational changes based on the collected data according to the physical process [28], [78], [79]. The PINN model exploits the data and the physical knowledge to guide the model [76]. The physical constraints are directly integrated into the initial loss function by penalizing the deviations from the target values. The use of PINNs allows identification to be performed simultaneously with the fitting of the FCNN model to a dataset generated with different parameters and conditions. This model exploits the automatic differentiation to generate every differential operator. Moreover, the model considers and captures the complex phenomena of structural locals void defects and deformations which characterize the quality of FSW welding. Unlike the physics-driven approaches, the hybrid approach helps to better understand the system of the physical relationship between the different components of the process. In addition to their computational performance, PINNs prove to be more robust, accurate, and generalizable than the traditional AI model. However, the PINN can suffer from several computational problems, mainly for the parameter estimation of multi-scale processes and the convergence of the loss function. Generally, the minimization of the total loss function represents a major issue for the PINNs including the traditional ANNs. Moreover, when the model is too complex (i.e., neural networks having several hidden layers) the approach can be inefficient and the estimated parameters may be ineffective for the investigated problem. Table 4.1 shows the PINNs approaches and their extension.

Table 4.1: Some applications of PINN models and their extensions.

Papers	Some applications of the PINNs and their variants
[291]	Active training of PINN to aggregate and interpolate parametric solutions to the NSE equations
[292]	PINN for the incompressible Navier-Stokes equations
[293]	Automating PINNs with error approximations
[294]	PINNs for fluid mechanics
[295]	PINNs for fluid mechanics Metamaterial Design
[296]	PINN for heat transfer problems
[297]	B-PINNs for forward and inverse PDE problems
[298]	Sparse PINN (SPINN) and interpretable NN for PDEs
[299]	Fractional PINN (FPINN)
[300]	Solving PDEs using DL and Physical Constraints
[301]	PINNs modeling of turbulent natural convection
[302]	Ultrasound computed tomography using PINN
[303]	Physics-informed model in wind turbine response prediction

4.3 Thermomechanical Process: Friction Stir Welding (FSW)

4.3.1 Principle and Operating Mode

Friction Stir Welding is a solid-state welding process that aims to assemble several similar or dissimilar materials with varying physical properties using a rotating and a translating tool mounted on the spindle of the machine [248]. This process is operated in several applications, and it is different from traditional welding processes like inert gas tungsten or laser welding. The materials to be joined are mechanically mixed, melted, and then solidified with a non-consumable rotating tool [248], [304]. The tool (see fig. 4.6) is composed of a shoulder that generates a major part of the heat for the welded materials softening and the pin which stirred the interface of the work-pieces. In fact, this piece is clamped on the bench of the welding machine. Furthermore, in the welding phase, the various parts to be joined are placed on the bench of a machine that has a forward movement in a vertical direction. The solid-state material flows occur due to interactions between the pin and shoulder workpieces. The authors of [305] show that the interaction between the tool and the part to be welded occurs in four main phases (see figure 4.3).

- Plunge Phase: During the first phase, the rotating tool is plugged directly into the joint line until the shoulder makes contact with the workpieces and plastics deform them.
- Dwelling phase: For this phase, the rotating tool stops translating in order to heat up the workpieces at a convenient temperature in the vicinity of the tool.
- Welding phase: In the welding phase, the rotating tool moves along the interface of the workpieces.
- Retracting phase: During this last phase, the material is driven periodically from the front to the back. At the end of the assembly process, the tool finally withdraws from the parts.

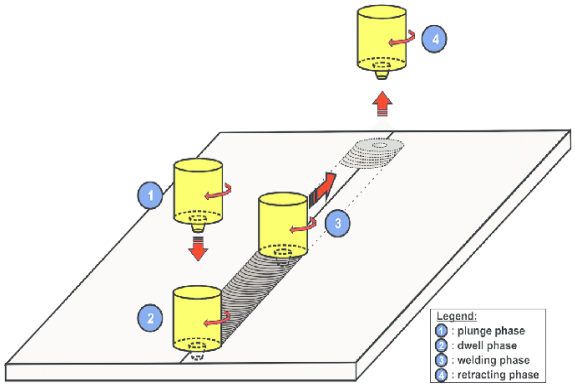


Figure 4.3: Scheme of the main phases of the thermomechanical Friction Stir Welding.

4.3.2 Parameters and Main Defects of the Process

This process is simple and highly efficient and is applicable to various industries. However, the mechanisms that govern the process and the deformation of the material around the tool can be very complex. Thus, the deformations in the welded area depend on several factors [241] such as the parameters of the welding process and the characteristics of the tools. In fact, some parameters can affect the quality of the resulting joints. We have mainly rotational velocity (rpm), welding velocity (mm/s), axial force (kn), tool shoulder diameter (mm), pin diameter (mm), pin length (mm), tool inclined angle (°), included angle of taper pin (°), pitch (mm), and the shoulder deepness inserted into the surface of base metal (mm) parameters. The authors of [240], [242] show that the defects are often dimensional (distortions and residual stresses) due to several factors including a poor calibration of the axial force, feed, rotation velocity, tool selection, and poorly retained parts. Paper [256] has demonstrated that the welding force keeps the tool pressed into the material, thus, the decrease of this force can lead to the training of a tunnel defect at the back of the pin. However, the high force can cause the tool to sink and the temperature of the material to rise. In addition to this, the defects due to the tool design are related to the incorrect combination of the tool and the workpiece to be welded or to the incorrect combination of the workpiece and the material of the support plate. Finally, we have other defects related to insufficient heat such as flash, surface seizure, vacuum bonding, kissing bonds, and onion rings.

4.4 Theories of the Numerical Simulation of the Thermomechanical Problem

4.4.1 Conditions of the model

In this study, we consider the model set up based on the work [255], including some conditions. We focus on all the phases of the process like the transition and fusion phases have a very short response time compared to the time of the whole process. We consider that the friction coefficient is constant, and we neglect the influence of the velocity of the friction coefficient. In addition, we select the plane in the middle of the part and the displacement is vertical. Also, to make high simulation accuracy, we need at least five layers of elements between two adjacent welding strips. Finally, we do not include the boundary conditions [306] in hybrid models.

4.4.2 Geometrical modeling

The numerical model is two-dimensional, but one can use three-dimensional models easily without any big changes. The foundations of two-dimensional modeling of FSW processes are detailed by [255]. The geometry of the model showing the tool and the model domain sizes are given in Figure 4.4. The model considers a small rectangular zone of size 12mm $\times$6

mm around the tool. As the welding tool pin is trigonal, $2D$ models plane crossing the pin exhibited and three curved sides of radius 5 mm whereas the tool maximal radius is 2.5 mm.

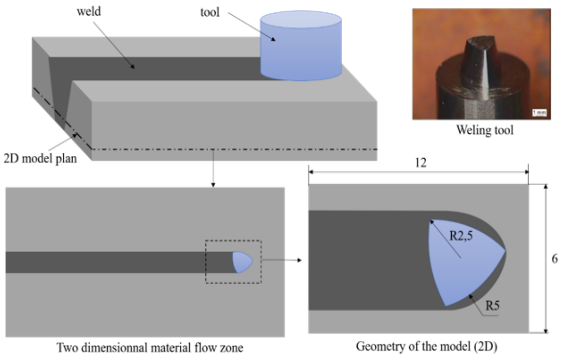


Figure 4.4: Model geometry of the thermomechanical Friction Stir Welding

4.4.3 Meshing and boundary conditions

The simulation of the friction stir welding process is a complex task, because during the non-stationary phase, several elements are involved, such as mechanics, heat transfer, or deformations around the tool axis. The meshed domain is Eulerian, the material is not fixed to the grid. As the meshed domain shape should fit the rotation tool shape, the blue zone of the mesh (see figure 4.5) rotates with the tool, and data are interpolated to the new position of the mesh at each time increment. This technique, known as " moving mesh ", has been used by [307] for CFD simulation of the FSW process. The model is compared to a Solid Mechanic FSW model in the work conducted by [308]. CFD method is an approach that helps to study the quantitative thermomechanical conditions of the fusion welding processes, such as the temperature and deformation field of the material. In the literature, several studies show that this approach is commonly used for modeling the FSW process [306].

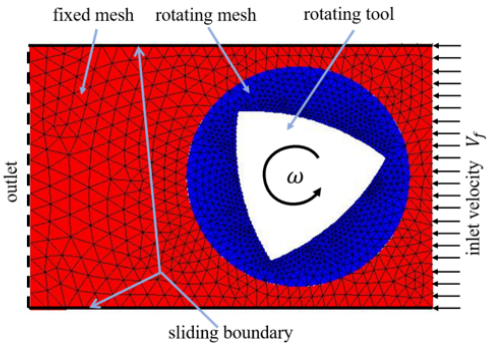


Figure 4.5: Model meshing and boundary conditions

The process requires that the finite element mesh is very fine to recover the history of the strong thermomechanical gradients undergone by the material. We are also interested in several elements such as fluid mechanics, the power of plastic deformation, and the "Finite Volumes" method with meshing based on the partial derivative equations of Navier-Stokes and Cauchy. The results of [309] show that to reproduce real welding behaviors concerning materials, temperatures, and deformation rates, the viscosity must be very low. Several studies have been realized in the boundary conditions [306]. In addition, we consider the following conditions: the tool does not mesh and its interaction with the material is modeled by applying sticking boundary conditions with shear limits of 300 MPa at the internal wall of the blue zone. The material enters the mesh domain at the inlet within the process feed velocity V_f. The sliding boundary condition is applied to the sides of the meshed domain whereas outlet pressure is set to atmospheric.

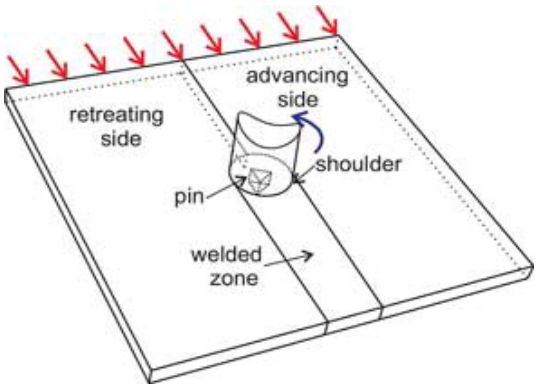


Figure 4.6: Diagram of the Friction Stir Welding process: The blue arrows represent the rotation velocity imposed on the pin, while the advancing velocity is replaced by a velocity imposed on the plates in the opposite direction (red arrows) [308].

4.4.4 Governing equations

The governing equations are represented by the PDEs, these equations are differential equations composed of multivariate functions and partial derivatives. PDEs can be applied in many areas, regarding the meteorology industries they can apply to charas, the flow of water in a pipe, and other phenomena involving the flow of various fluids. In the mechanics and energy sectors, NSE aims to design the means of transport including airplanes, automobiles, trains, and electrical power plants. In this case, the PDEs (incompressible NSE) describe the physical constraints [76]. A general form of the function $u(X,t)$ can be defined in the following form:

$$f\left(X,t,\hat{u},\partial_x\hat{u},\partial_t,...,\lambda\right)=0,\quad X\in\Omega,\quad t\in[0,T] \tag{4.1}$$

Where the function f describes the residual of the equation, $X = (x, y) \in \mathbb{R}^d$ is the network inputs or the features (partial coordinates are space and t is the time). The parameters of the PDE are described by the differential operators like $[\partial_x \hat{u}, \partial_t \hat{u}, ...]$; the parameters $\lambda = [\lambda_1, \lambda_1, .., \lambda_n]$ and their solution is given by $\hat{u}(X, t)$. In practice, this equation can be solved by an approximate solution using some numerical methods (FVM), or Deep Learning (DL). The material flow during the welding phases is calculated by solving the NSE and Cauchy partial differential equations under the constraint that the viscosity does not depend on the deformation rate. Also, the governing equations can be derived from Cauchy momentum equations under some assumptions. The general convective form of Cauchy momentum equations is given by:

$$\overrightarrow{\nabla} \cdot \overline{\overline{\sigma}} + \rho \overrightarrow{\mathrm{g}} = \rho \frac{D\overrightarrow{v}}{Dt} \tag{4.2}$$

Where $\overline{\overline{\sigma}}$ is the Cauchy stress tensor, ρ the density, $\overrightarrow{\mathrm{g}}$ the body accelerations (gravity, inertial accelerations, or electrostatic acceleration), $\overrightarrow{v}$ the material flow velocity and t the time. Using $\overline{\overline{\sigma}} = \overline{\overline{\tau}} - p\overline{\overline{I}}$, where $\overline{\overline{\tau}}$ is the deviation stress, p the hydrostatic pressure and $\overline{\overline{I}}$ the second order identity tensor, equation (4.2) can be rewritten as follows:

$$\overrightarrow{\nabla} \cdot \overline{\overline{\tau}} - \overrightarrow{\nabla} p + \rho \overrightarrow{\mathrm{g}} = \rho \frac{D\overrightarrow{v}}{Dt} \tag{4.3}$$

$$\overrightarrow{\nabla} \cdot \overline{\overline{\tau}} - \overrightarrow{\nabla} p + \rho \overrightarrow{\mathrm{g}} = \rho \frac{\partial \overrightarrow{v}}{\partial t} + \frac{D\overrightarrow{v}}{Dt} \tag{4.4}$$

In addition, the mass conservation, or the equilibrium condition of the NSE equation will be assumed:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \overrightarrow{v}) = 0 \tag{4.5}$$

Furthermore, we consider that the gravity $g = 0$, and by assuming mass conservation, we obtained the conservative form of Cauchy stress:

$$\frac{\partial}{\partial t}(\rho \overrightarrow{v}) + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{v} \otimes \overrightarrow{v}) = -\overrightarrow{\nabla} p + \nabla \ \overline{\overline{\tau}} + \rho \overrightarrow{\mathrm{g}} \tag{4.6}$$

In general, for modeling Severe Plastic Deformation (SPD) process as FSW with CFD, incompressible Navier-Stokes equations are used as plastic deformation is incompressible. Considering incompressibility, the conservative Cauchy stress is given by the equation (4.6). The Navier Stokes Equation in a vector form is as follows:

$$\rho \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{t}} + \rho(\overrightarrow{v}\nabla)\overrightarrow{v} = \overrightarrow{F} - \nabla P + \frac{\mu}{\rho} \nabla^2 \overrightarrow{v} \tag{4.7}$$

4.4.5 Parameters and Mechanical properties

Table 4.2 shows the mechanical properties of the material for the numerical model simulation. The material properties have been defined by a coefficient k, the behavior law is made by the coefficients strip thickness n, the viscosity μ is defined by the relation $\mu = K\gamma^{n-1}Pa$ the density is ρ and the tool radius is R.

Table 4.2: Mechanical properties of material

Variables	**Description**	**Values**
U	Rotation velocity	2000 rev/ min
V	Velocity of advance	600 mm/min
N	Number of mesh nodes	825 nodes
T	Time step simulation	0.0125 sec
k	Consistency	1.5 $e^8 kg.sm^{-2}$
n	Strip thickness	0.014
ρ	Density	$2710 kg.m^{-3}$
μ_{min}	Minimum viscosity	$1000 kg/m/s$
μ_{max}	Maximum viscosity	$5e^{10} kg/m/s$
ϵ_{min}	-	0.01
$shear$	Shear limit at the tool material	$300e^{10} kg/m/s$
R	Tool radius	2 mm
L	Length field	12 mm
l	Width field	6 mm

4.5 Proposed Methodology

In this work, we propose a framework that has several objectives. The developed model addresses a solid mechanics problem where the viscosity is NSE. It aims to learn the process of the transient regime, furthermore, it predicts the parameters during the whole process (transient and stationary regime). PINNs have several advantages such as (a) the choice of network architecture since we can impose governing equations on the FCNN inputs, thus having a considerable impact on the FCNN outputs; (b) Exploiting sophisticated automatic differentiation algorithms for accurate differentiation of FCNN functionals and for back-propagation of errors. (c) Exploitation of advanced machine learning software with parallel processing capacities by CPU and GPU, and TensorFlow and Pytorch. Figure 4.7 describes the workflow for the proposed methodology that is based on two main modules. This graphic is composed

of several blocks. The first module focuses on the numerical simulation of the FSW Process. This block aims to generate real by using the FVM method and commercial code Ansys/Fluent. For the simulation, we fixed values on some parameters or the mechanical properties of the tool (see table 4.2). The second block represents the result of the simulation. The output data is a set of spatial coordinates, velocities, pressure, and time. The third block consists of analyzing the data and providing a summary or descriptive statistics. This analysis helps to improve the insight and identify any missing values or potential outliers in the data. The fourth block performs some pre-processing techniques, such as the Min-Max normalization (see equation 4.8) and the re-sampling technique. Normalization is the process of eliminating the effect of size between data or changing the values in the data set to use a common scale, without distorting them.

The fifth block is represented by figure 4.10, this is the key contribution of this chapter. The PINN model is an extension of the FCNN that does not require the use of high-fidelity of simulated data. Thus, the model optimal helps to identify, and predict several parameters of the process. In fact, to estimate the optimal parameters, we exploit the physical prior knowledge in the form of PDEs to regularize the loss function. The PDEs are computed using automatic differentiation. The regularization technique improves the generalization performance and guarantees that the solutions are consistent with the physics laws. By training an FCNN we are interested in optimizing the combined loss function (see equation 4.21). We will provide more details when presenting the Schematic of PINN. The last block consists of visualizing, validating, and interpreting the estimated values and parameters. We, therefore, validate the model through the RMSE.

4.5.1 Data Pre-Processing

To train the developed model, we performed several pre-processing operations including the data Min-Max Normalization. This operator is used to limit the size effect of each feature. The min-max normalization is represented by the following equation:

$$X_{Norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad where \;\; X_{Norm} \in (0, 1) \tag{4.8}$$

4.5.2 Type of Layer Activation Functions of the Feed-forward Neutral Networks

In this section, we provide an answer to the following questions: *why does FCNN need a specific activation function? How to choose the right activation function?* To answer these questions, we introduce the formal neuron (see figure 4.8) which is an algebraic parametric and nonlinear function with bound values. The activation functions have an important contribution to the training of FCNNs, in fact, the function assists to learn the input information and making sense of the non-linear mappings between the neurons' inputs and outputs. More-

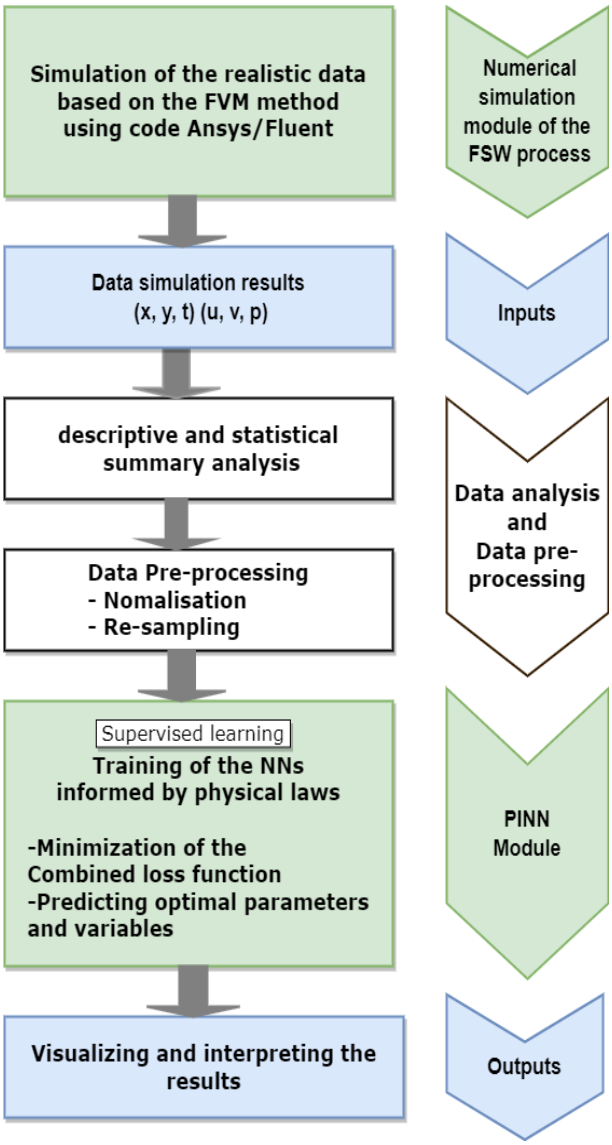


Figure 4.7: Detailed flow chart of the developed methodology. This methodology is mainly based on two modules, namely the numerical simulation module and the PINN approach applied to 2D data.

over, the activation functions help to make a dynamic network, thus improving the ability to extract complex information from data generated by nonlinear systems. In other words, the activation functions are exploited to transform an input into an output signal. This signal is then used as an input signal to the next layer. Furthermore, the function brings the nonlinearity of the system. The inputs x and outputs y of the activation function respectively are presented by the following equations:

$$x = \varphi\left(w_0 + \sum_{j=1}^{n} x_j w_j\right) = \varphi \sum_{j=0}^{n} x_j w_j \tag{4.9}$$

$$y = f(x) = f\left(\varphi \sum_{j=0}^{n} x_j w_j\right) \tag{4.10}$$

where w_j, is the neuron's synaptic weight matrix, w_0 the bias vector of 0 input set to 1, and φ describes a type of function applied to an artificial neuron's output function.

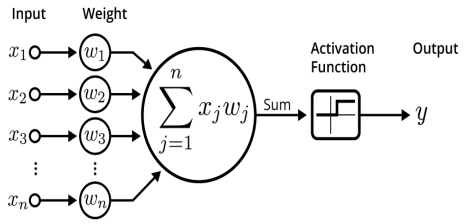


Figure 4.8: Formal neuron represented by x_n inputs, output y and a given activation function.

In the literature, there are various types of activation functions that are commonly used in neural network approaches [310]. The choice and the property of this function can significantly impact or influence the precision result of the model. When we use a linear function like "Identity", the global computation performed by the network will also be linear, then it will be useless to use several neurons because a single neuron will give equivalent results. If the function is polynomial, this may increase the computational time necessary to estimate the model parameters. However, in some contexts, it is also important to ensure that the network outputs are not limited or bounded. In this case, it is preferable to use a "linear" function instead of a "Sigmoid" type function which is bounded (it helps to accelerate the derivative computation time and to reduce the computational time required to train the neural network) [311]. Furthermore, for the "Piecewise Linear" type, we can show that these functions do not have a certain form of regularity and it not differentiable on singular points. In fact, it is not possible to compute the gradient of the error of the model on these points. The regularity of the activation function favors the learning because to compute the Hessian matrix, it is necessary that the function is twice derivable. The most used functions are mentioned in the following list:

- Identity: $\varphi(x) = x$
- Logistic, Sigmoid: $\varphi(x) = \frac{1}{1+e^{-x}}$.
- Hyperbolic tangent (Tanh) $\varphi(x) = tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- Function Sign: $\varphi(x) = \frac{x}{1-|x|}$

- Rectified Linear Unit (ReLU): $\varphi(x) = max\{0, x\}$
- Leaky ReLU: $\varphi(x) = 0.01x$ if $x < 0$ and x if $x \geq 0$
- Gaussian: $\varphi(x) = e^{x^2}$
- Softmax: $\varphi(x) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}$, where $j = 1, 2, ..., J$

In this case study, we focus on the ReLU function [312]. This monotonous function is very often used in NN for its simplicity and ability to capture interactions and non-linearity. In fact, Relu is the most common choice for feed-forward neural network regression because it is not subject to the explosiveness or disappearance of the gradients [313] and their derivative is equal to 1. Sometimes, the Leakage ReLU function is used in the context of functions with zero derivatives. In addition, ReLU is applied to the hidden layers to facilitate the convergence of the gradient descent and thus, encourage faster learning compared to the Sigmoid and Tanh functions. Moreover, it can learn reliably even when the number of layers increases. The aim of this function consists of enabling the neuron set to be excited. Then the inputs are multiplied by weights and the bias is added to the sum of the product obtained. The result is then converted into a signal indicating the state of neuron excitation. The ReLU function must be used only in the hidden layers and not the outer layer. Furthermore, the hidden layers help to extract features from the input x_i data using the following equation:

$$h_j = RELU\left(\sum_{i=0}^{n} x_i w_{ij} + b_i\right) \tag{4.11}$$

where y_j is the j^{th} output data, w_{ij} is the kernel, b_j is the bias coefficient, the coefficient x and b are biases updated in the same way. and x_i is a feature of previous layers. The output vectors for the k^{th} hidden is given by:

$$h_i^k = ReLU\left(\sum_j w_{ij}^{(k-1)} h_j^{(k-1)} + b_i^{(k)}\right)$$

where $k = 1, 2..., h^{(k)}$ is the output of the k^{th} layer, w_{ij}^k and $b_i^{(k)}$ are respectively the weight and the bias of the k^{th} layer.

4.5.3 Regularization and Physics-based Loss function

To train the model and to estimate the optimal parameters we use the common optimizer like the Stochastic Gradient Descent (SGD) [314]. In this case, we focus on the Adam algorithm that helps to find an optimal solution to the learning problem. Furthermore, Adam allows to gradually correct the parameters in order to minimize a continuous and differential function named Loss function ($\mathcal{L}$). Along the same lines, the network learns to approximate the

differential equations by finding the parameter θ or to estimate a predictor function $\hat{Y}$ through the minimization of the loss function. The equations of the traditional loss function $\mathcal{L}(\theta)$ and the optimization problem $\mathcal{P}$ are respectively defined by the following:

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{1}^{n} l(f(x_i; \theta), y_i) \tag{4.12}$$

$$\mathcal{P} : \theta = arg \min_{\theta} \mathcal{L}(\theta) \tag{4.13}$$

Where $f(x_i; \theta)$ describes a given model. Despite their advantages, the FCNN-based models are often subject to the overfitting problem in the training stage, which can result in a high variance in the test data. To address this issue and obtain a more robust and generalizable model (accuracy, simplicity, and consistency), we develop the PINN models, that exploit a fully connected feed-forward neural network, which is formed by multiple hidden layers. This approach is based on the combined loss function, in addition, it is a new contribution compared to the classical models. The regularization methods consist of penalizing the weights of neurons; thus, the resulting model must have an optimal prediction (minimum error) on all sets of training data. In this respect, several approaches have been developed to fit the model and then find a compromise between the bias and the variance. The first method is to determine the optimal number of hidden layers and neurons. However, the author of [315] demonstrates that this theoretical method to reduce the number of neurons during training has limitations mainly when the training data includes few samples. By investigating different architectures for solving regression problems, the bias and variance do not necessarily evolve in opposite directions when the number of hidden neurons increases or decreases. An alternative approach is to investigate the possibilities of obtaining estimators with reduced amplitudes, i.e., to accept a slightly increased bias in order to reduce the variance more than proportionally. Regularization provides this type of insight, furthermore, it is one of the key techniques of the AI framework that aims to limit the overfitting issue and achieve better performance. There are several regularization techniques classified as passive or active methods.

4.5.3.1 Passive Regularization techniques based on data

The most popular passive method consists of using a validation database to evaluate the performance of the model. This approach is efficient for regression tasks because the network tends to fit the data on the whole space. Furthermore, the variation of the loss function on the validation database is easier to detect. If the model is not adjusted excessively to the data, the loss function associated with the validation and learning databases decreases together. When the generalization error increases, the cost function on the learning base continues to decrease, while the validation one tends to increase. However, this method requires a high amount of data to train, test, and validate the model.

4.5.3.2 Active Regularization techniques based on data

To address the challenges of the passive method, we introduce the most important used active regularization approaches in the AI framework. Early Stopping is an approach that consists of stopping the iterations of the optimization algorithm before it convergences. If this convergence is not achieved, the model does not fit the training data properly, thus limiting the overfitting effects. To apply this method, it is necessary to determine the optimal number of iterations to be used during the training step. So, a basic method consists of following the cost function evolution on a validation basis and stopping the iterations when the cost calculated on this basis is increasing. However, the early stopping method can be inapplicable, in some situations, because it can be difficult to determine with exactitude the moment to stop the training since the performances on the validation base does not degrade significantly. Penalizing the parameter weight technique is based on the limitation of the model's capacity, by adding a parametric norm penalty to the loss function [261], [316], [317]. There are different types of regularization according to the norm or distance metric applied to the weight. (i) The Weight Decay is based on the L_2 norm parameters like the Ridge regression model, the function is described by the following relation:

$$\mathcal{L}(\theta) = \frac{1}{2}||w||_2^2 = \frac{1}{2}\sum_{i=1}^{n} w_i^2 \tag{4.14}$$

When the weights of the network are high in absolute value, the hidden neurons' sigmoid is saturated, so the modeled functions can have abrupt variations. To obtain regular functions, we need to work with the linear part of the sigmoid, which implies having weights whose absolute value is small [318]. We can add a penalty to the cost function that depends on the magnitude of the weights that link the neurons together. Thus, the weights w_i are approximated towards the origin by adding the regularization term in the loss function. (ii) The Sparse representation has the same properties as the LASSO regression method. This approach is based on the L_1 norm that directly penalizes the activation of a neuron, instead of its weight. Thus, the sparse representation performs the feature selection task which produces sparse solutions whose weights w_i are set to 0 for λ large enough. The Sparse representation is defined by the following relation:

$$\mathcal{L}(\theta) = \frac{1}{2}||w||_1^1 = \sum_{i=1}^{n} |w_i| \tag{4.15}$$

However, this approach has its limitation since the derivative form is unknown, so it is based on an approximation of the derivation. (iii) The dropout regularization is based on the constraint of the maximum norm, but it is limited to the absolute norm of the neuron weights ($||w||_2 < c$ where $c > 0$). This method prevents the neural network from "exploding" and anticipates the problem of overfitting [319].

The term "Dropout" refers to the temporary suppression of the neurons (both hidden

and visible neurons). Figure 4.9 shows that the neural network is randomly cut off a part of its neurons during all the n iterations of the training phase. However, during the model validation phase, these neurons are reactivated (p factors). The dropout is more robust than other regularization methods such as $L2$ regularization since the dropout training process helps to explore different regions of the parameter space that it would not have found during a regular training [319]. In the general case, the dropout method applied to the units produces the following random variables:

$$S_i^h = \sum_{i<h} \sum_j \delta_i^l w_{ij}^{hI} . S_j^l \tag{4.16}$$

By assuming that the dropout process is independent of the activities of the weights, when the dropout is applied to the units, we obtain the following relation:

$$\mathbf{E}(S_i^h) = S_i^h = \sum_{i<h} \sum_j \delta_i^l w_{ij}^{hI} . \mathbf{E}(S_j^l) \quad for \quad h > 0 \tag{4.17}$$

where $S_j^0 = I_j$ and δ_i^l is the random variable of the type Bernoulli selector, which suppresses the influence of weight w_i with the probability $P(\delta_i = 0) = q_i$. The equation (4.17) can be applied recursively (backpropagation) to the entire network, including the input layer. Furthermore, by assuming that the Bernoulli random variables $(\delta_i^l, \delta_{i'}^{l'})$ are independent of $(S_i^l, S_{i'}^{l'})$, when $i \neq i'$ and $l \neq l'$, we deduce that the probabilities $P(\delta_i = 1) = 1 - q_i = p_i$, $\mathbf{E}(\delta_i^l, \delta_{j'}^{l'}) = P_j^l, P_{j'}^{l'}$ and $\mathbf{E}(\delta_j^l, \delta_j^l) = P_j^l$. Finally, we obtain the following equation:

$$\mathbf{E}(S_i^h) = S_i^h = \sum_{i<h} \sum_j P_{ij}^{hl} w_{ij}^{hI} . \mathbf{E}(S_j^l) \quad for \quad h > 0 \tag{4.18}$$

In order to obtain models that are not overfitted the authors of [320] have shown that the value of the weights is more important than their number.

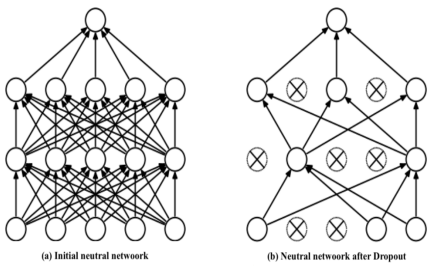


Figure 4.9: Dropout regularization method applied to the weight of the Neural Network. Figure (a) shows the initial state of the network, and figure (b) indicates the momentary suppression of the neurons during the network training.

4.5.3.3 Regularization Approaches based on the Physical Constraints

Let's introduce the functions f and $\mathcal{F}$ that respectively represent the function that determines the problem data and the parametrized PDE equation expressed in the most general system as:

$$\mathcal{F}(u(X,t);\lambda) = f(X,t) \quad X \ in \ \Omega, \quad t \ \in \ [0,T] \tag{4.19}$$

The model $\mathcal{F}(u,p)$ is defined in the spatial domain $\Omega \subset \mathbb{R}$. By assuming $\mathcal{F}(u,p) = 0$, we obtain the following equation:

$$\begin{cases} \nabla.u = 0 \\ \frac{\partial u}{\partial t} + (u.\nabla)u + \frac{1}{\rho}\nabla p - v\nabla^2 u + b_f = 0 \end{cases} \tag{4.20}$$

where $\mathcal{F}(u,p) = f(\hat{u}, \partial_t\hat{u}, \partial_x\hat{u}, ..., \lambda) = 0$, $X \in \mathbb{R}^d$ represents ,the spacial coordinate t is the time. f describes the residual of the PDE containing the differential operator such as $\hat{u}, \partial_t\hat{u}, \partial_x\hat{u}, ...$, and $\lambda = [\lambda_1, \lambda_2...,] \in \mathbb{R}^d$ are the PDE parameters. The aim of this problem is to obtain the function u for each z, with parameters λ that modify the weight of the constraints, the solution of the equation is described by $\hat{u}(X,t)$. We enforce explicit regularization $\mathcal{L}_\mathcal{F}(\theta)$ terms to provide effective control and insight into the complexity of the PINN model at the fine-tuning stage. Furthermore, this approach consists of driving the model to follow the physical process or topology. The regularization function related to the physical constraints penalizes the residual of the governing equations and it is defined by the following relation:

$$\mathcal{L}_\mathcal{F}(\theta) = f(\hat{u}, \partial_t\hat{u}, \partial_x\hat{u}, ..., \lambda)$$

By adding the dropout regularization and the Physical constraints terms to the initial loss function (see equation 4.13), we obtain the total loss function $\mathcal{L}(\theta)$ which is a linear combination of some losses from data and PDE residuals.

$$\mathcal{L}(\theta) = \lambda_d\mathcal{L}_{data}(\theta) + \lambda_D\mathcal{L}_{Dropout}(\theta) + \lambda_\mathcal{F}\mathcal{L}_\mathcal{F}(\theta) \tag{4.21}$$

In this context, the network must learn to approximate the PDEs by finding the parameter θ that minimizes the loss function. The equation of the estimated parameter $\hat{\theta}$ is:

$$\hat{\theta} = arg\min_\theta(\lambda_d\mathcal{L}_{data}(\theta) + \lambda_D\mathcal{L}_{Dropout}(\theta) + \lambda_\mathcal{F}\mathcal{L}_\mathcal{F}(\theta)) \tag{4.22}$$

We note that the function $\mathcal{L}_{data}(\theta)$ concerns the errors in the description of the initial state loss function. $\mathcal{L}_{\mathcal{F}(\theta)}$ is the loss function produced by a mismatch with the governing differential equations (NSE). This function applies the differential equation $\mathcal{F}$ to the collocation points,

which can be chosen on the domain Ω. In addition, by using the PDE constraints through the penalty term $\mathcal{L}_{\mathcal{F}}(\theta)$, the corresponding weight $\lambda_{\mathcal{F}}$ can be applied to address the accuracy of the PDE model. Furthermore, the hyper-parameters λ_d, λ_D and $\lambda_{\mathcal{F}}$ are to be tuned and their purposes are to influence or measure the importance of the regularization function weights. The penalty coefficient related to the data $\mathcal{L}_d(\theta)$ denotes the error based on the observation data or the validation of known data points. The coefficient penalty $\lambda_D(\theta)$ is the term related to the regularization of the model's parameter. The coefficient of the Physical constraints $\lambda_{\mathcal{F}}$ is the term that fits the effects of the physics inconsistency on the empirical loss function and the model complexity. By considering the weight $\lambda_{\mathcal{F}} = 0$, we lose knowledge about the physical process. In this respect, the traditional FCNNs models are trained without any insight into the dynamics of the physical phenomenon. We can therefore infer that the penalty terms aim to guide the model, in order to follow the physical structure of the process.

4.5.4 Observations about the Combined Loss Function

For our case study, the PINN model is applied as a supervised learning task. The PINN framework has $(X,t)^N = \{(x_i, y_i, t_i)\}_{i=1}^N$ as the inputs and $Y^N = \{(u_i, v_i, p_i)\}_{i=1}^N$ as the output variable where u_i, v_i, and p_i respectively provide descriptions of the velocities and total pressure. The optimization problem described by the equation 4.13 can be addressed by applying the algorithm Adam, which minimizes the combined loss function. On the other hand, by using this Adam, the parameters of the model are estimated by minimizing the difference between the observed outputs and the model's predictions. This minimization is performed in several steps: (a) The spatial coordinates of the collocation points and the training data are substituted into the loss function. (b) The spatial and temporal derivatives with respect to weight and bias are performed on the loss function. These derivatives are accurately and efficiently calculated using automatic differentiation (AD). The AD does not suffer from truncation or rounding errors, which results in much higher accuracy. Furthermore, the algorithm avoids bad local minima and improves the speed of convergence it combines adaptive learning rate and momentum methods. (c) Using the gradient descent to update the w and b vector. The values $\hat{y}_i$ are the variable to be predicted for the value of i^{th} the index of a training example, y_i the true value of the i^{th} samples. For the purpose of regression, there are several forms of the loss function to evaluate the performance of the NNs. We can mention the Root Mean Squared Error, Squared Error, Absolute Error, and Huber Error. In this case study, we use a Root Mean Squared Error that has the following form:

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{N}(y_i - \hat{y}_i)^2} \tag{4.23}$$

where $\hat{y}_i = (\hat{u}_i, \hat{v}_i, \hat{p}_i)$ respectively are the vectors of the target and variable to be predicted. To obtain the optimal parameters of the PINN, we minimize the following mean square error:

$$RMSE_{(total)} = RMSE_{(data)} + RMSE_{(dropout)} + RMSE_{(\mathcal{F})} \tag{4.24}$$

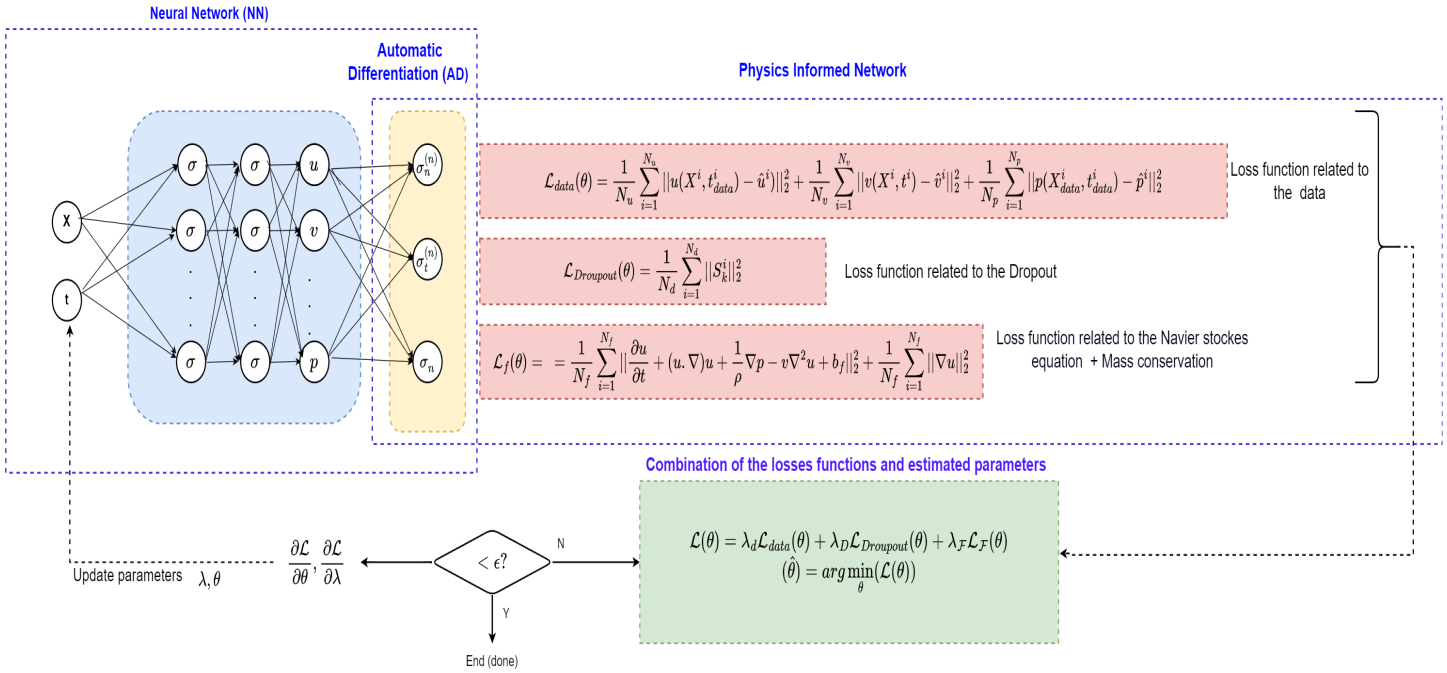


Figure 4.10: Schematic of physics-informed neural network (PINN). This schematic represents a detailed description of block five of the figure 4.7). PINNs are composed of residues (losses) of differential equations 4.21. The FCNN takes as inputs the vector $(X,t) = [(x,y),t]$ where (x,y) and t are respectively the spacial coordinates and times. This network is trained to approximate the physics solutions $\hat{y} = [\hat{u}, \hat{v}, \hat{p}]$. We used Automatic Differentiation (AD) to compute the derivatives of parameters with respect to the input vector. Then AD helps to express the residuals of the governing equations in the combined loss function. This function is generally composed of several terms weighted by different coefficients. The last step containing all residuals is known as the feedback mechanism. This aims to minimize the loss function, using the Adam optimizer, based on some learning rate in order to obtain the parameters θ of the network. In fact, the parameter θ of the FCNN and the parameter λ of the unknown PDE can be estimated simultaneously by minimizing the combined loss function.

where $RMSE_{(total)}$ represents the combined mean square error generated by the developed PINN model and each term is written as:

- $RMSE_{(data)} = \mathcal{L}_{data}$ is a sum-of-square error function between the target output, y_i and the predicted output $\hat{y}_i$. This function is based on the observation data of the initial condition.

$$\begin{aligned} RMSE_{(data)} = & \frac{1}{N_u}\sum_{i=1}^{N_u} ||u(X^i,t^i) - \hat{u}^i)||_2^2 \\ & + \frac{1}{N_v}\sum_{i=1}^{N_v} ||v(X^i,t^i) - \hat{v}^i||_2^2 \\ & + \frac{1}{N_p}\sum_{i=1}^{N_v} ||p(X^i,t^i) - \hat{p}^i||_2^2 \end{aligned}$$

- $RMSE_{(dropout)} = \mathcal{L}_{Dropout}(\theta) = \frac{1}{N_d}\sum_{i=1}^{N_d} ||S_k^i||_2^2$

- $RMSE_{(\mathcal{F})} = \mathcal{L}_{\mathcal{F}}(\theta)$ is the function based on the physical constraints of the PDE and mass conservation.

$RMSE_{(\mathcal{F})} = \frac{1}{N_{\mathcal{F}}}\sum_{i=1}^{N_{\mathcal{F}}} ||\mathcal{F}(u,p)||_2^2 = \frac{1}{N_{\mathcal{F}}}\sum_{i=1}^{N_{\mathcal{F}}} ||\frac{\partial u}{\partial t} + (u.\nabla)u + \frac{1}{\rho}\nabla p - v\nabla^2 u + b_f||_2^2 + \frac{1}{N_{\mathcal{F}}}\sum_{i=1}^{N_{\mathcal{F}}} ||\nabla u||_2^2$

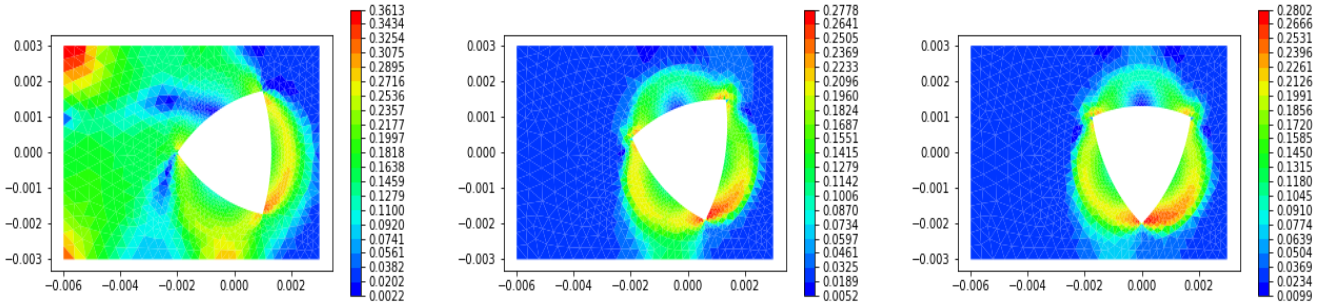
where $RMSE_{(\mathcal{F})}$ is the N-S equation, and the Mass conservation. We can notice that $N_u = N_v = N_d$ and $N_{\mathcal{F}}$ are the number of training examples in the training set. To minimize the cost function (see equation 4.24), we exploit an adaptive gradient-based algorithm named Adaptive Moment Estimation (Adam) which facilitates SGD convergence by combining both an adaptive learning rate and a momentum method [321]. In other words, the PINNs models, solve a PDE system (See equation 4.21), and the model converts the equation into an optimization problem by updating iteratively the parameter θ in order to minimize the final loss function. To obtain a higher learning rate and to estimate the optimal hyper-parameters including the model regularization coefficients we investigate an optimal FCNN architecture.

Figure 4.10 shows the zoom-in or more detailed information of block five of figure 4.7. Specifically, figure 4.10 represents the schematic of a physics-informed neural network. This framework is the most common problem in fluid mechanics, where the PDE considered in our application is the approximation of the nonlinear NSE. In addition, minimizing the combined loss function, also allows us to identify and estimate the optimal parameters of the process. The key components are respectively the Fully-Connected Neural Network (FCNN), Automatic Differentiation (AD), the governing equations, and the combined loss function (see equation 4.21). By using the chain rule, the AD can compute the partial derivatives of the outputs according to the inputs directly in the computational graph via a combination of the derivatives based on a sequence of operations. Unlike conventional numerical methods, the AD approach performs the computation of the partial derivatives via an explicit expression, thus avoiding the introduction of discretization and truncation errors. However, we can note that there are generalization and optimization errors that depend respectively on the training data and the optimizer used.

4.6 Results and Discussions

4.6.1 Result of the Numerical Simulation of the 2D Data

In this subsection, we present the main results of the developed framework including the numerical solution of the FSW process and the PINN module. The first step consists to simulated the $2D$ data using the FVM approach and CFD software Fluent. We have considered all the phases of the thermomechanical process. In addition, we have a two-dimensional model with plane stress of the union between two plates. Our model is composed of 850 nodes, furthermore, we assume that the axis rotates at 2000 rev/min and advances at $600mm/sec$. The time increment is 0.0125 sec, the strip thickness is equal to 0.014, and the tool radius, length field, and width field are respectively equal to $2mm$, $12mm$, and $6mm$. In addition, we consider the geometric parameters defined in the sub-section 4.4.2. These data are represented by the the spatial coordinate (x, y), time step t, $2 - zone$, the velocities (v_x, v_y) and the total pressure (p). We have randomly chosen some points that correspond to time steps t equal to 0.0060, 0.0765, and 0.150 sec. Figures 4.11 represent the contours plot or evolution of the components (velocities, and pressure) at different time steps. Specifically, the plots (sub-figures 4.11 (a), (b), and (c)) represent respectively the evolution of the velocities, and the pressure field during the beginning, middle, and end of the process.



(a) (b) (c)

Figure 4.11: Simulation results of the $2D$ process (Contour plots): Contour plots represent the evolution of the velocities and pressure subsequently at the beginning, middle, and end of the process. For each step, the time steps are respectively equal to 0.0060, 0.0765, and 0.150 sec).

4.6.2 Physics-Informed Neural Networks Trained on the FVM Solution

In the second step, we train and validate the PINN model of the thermal-fluid model that uses a Fully Connected Neural Network (FCNN). The network is governed by physic law imposed in the loss function. The model is a valid substitute for the numerical solution.

The purpose of the hybrid approach is to perform both prediction and validation of the simulated data. To illustrate the application of the proposed approach, we use the simulated synthetic data to estimate the hyper-parameters and predict the output data. Furthermore, the model is tested and tuned through different scenarios or configurations (see table 4.3). This provides some highlights on the performance of the proposed method. The robustness of the model is influenced by various parameters. We can mention the hidden layers, the number of neurons, the training methods, the activation function, batch size, the number of epochs, the optimization algorithm, and the loss function. PINN modules can reduce the approximation error while increasing the trainability of the network. In addition, it can also provide a large generalization error, thus hyperparameters, such as the learning rate, or the number of iterations can be adjusted to control and improve the overfitting problem. Concerning the development of the model, we select the training data randomly from the generated date set, furthermore, we exploit a fully connected neural network with some hidden layers. To activate each layer, we use the Relu function, in the fine-tuning phase we operate with the Adam optimizer to perform the optimization up to convergence. This, algorithm help to ensure global convergence and accelerate the process of convergence. Subsequently, the optimal parameters θ and λ are computed by minimizing the sum of the loss function (see equation 4.24) iteratively until it satisfies the stop criteria.

Table 4.3: Variations of the hyper-parameters of the PINN model

Parameters	**Different values**
Learning rate	$1e^{1}, 1e^{-1}, 1e^{-2}, 1e^{-3}, 1e^{-4}, 1e^{-5}$
Training data (%)	50, 55, 60, 65, 70, 75, 80, 85, 90, 95
Number of Epochs	50, 100, 150, 200, 250, 300, 350 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000
Number of Layers	4, 8, 12, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 40, 50
Number of neurons	2, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60
Dropout coefficient	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1

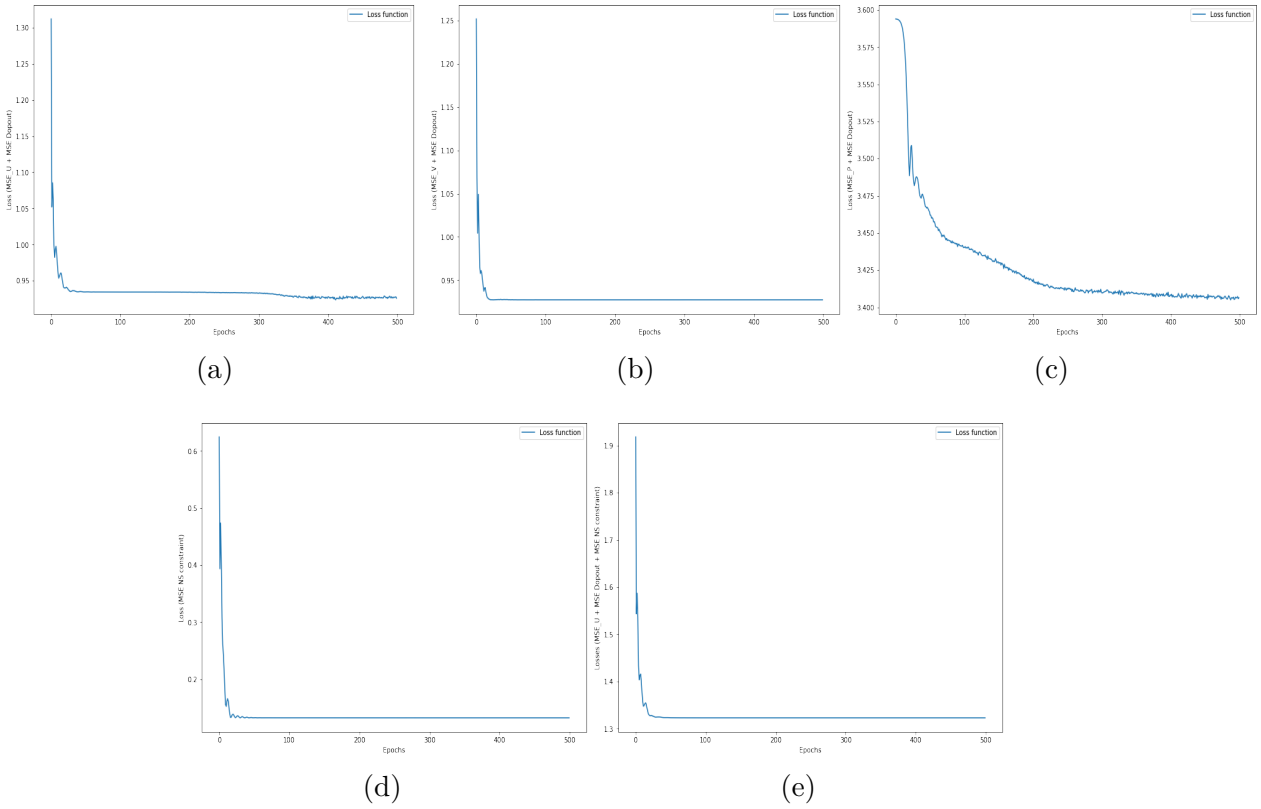


Figure 4.12: Losses functions: Each figure (a), (b), (c) and (d) respectively represents the loss function related to the variable u, v, p and the derivatives for the NSE. Furthermore, sub-figure (e) show the combined loss function defined by the equation 4.21, with respect to the number of training epochs

Figures 4.12 (a), (b), (c), and (d) show the learning performance (loss) curves of the algorithm as a function of the number of epochs. These curves can be regarded as a potential tool for diagnosis and to evaluate the quality of the training algorithm when updating the parameters. In fact, each figure shows respectively the loss value relative to the output data and physical constraint like NSE. We can deduce that the loss functions related to the variable (u, v) and the physical constraints (NSE) have a fast decay. Concerning the loss function related to the variable p, we have applied the logarithm to favor or accelerate the convergence of the function. sub-Figure 4.12 (e) shows the combined loss function including the dropout regularization function that converges much faster and more regularly. The performance obtained with this novel approach is remarkable and the total loss is significantly lower when the training is completed. The selection of the optimal network is performed according to the lowest value of the RMSE. Thus, the architecture of the optimal model is composed of 8 hidden layers, and 5 neurons by layers. In addition, the network is trained during 1000 epochs with a learning rate equal to 0.1 and the coefficient of dropout regularization is 0.3. The RMSE computed through the optimal model is 1.93, when we remove the variable p from the model the RMSE decreases this value decrease to 5.84. Despite the logarithmic

transformation of the loss function, the model suffers in the presence of this variable. Figure 4.13 represents the contour plots of the $2D$ FSW process at the beginning of the process. Sub-figures 4.13 (a, and b) represent the real data simulation, sub-figures (c), and (d) show the predicted values obtained with test data equal to 5% at different times steps.

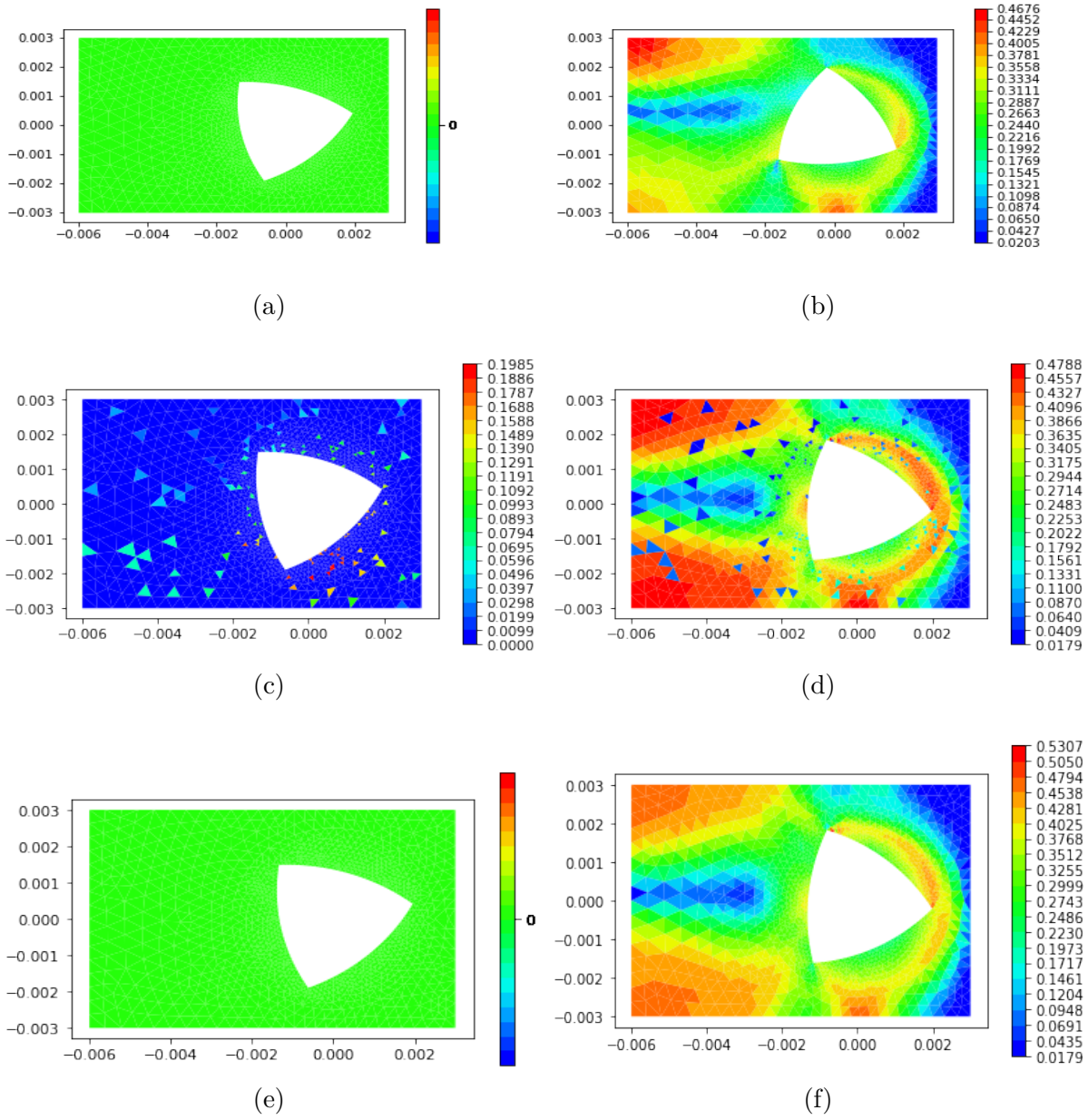


Figure 4.13: A comparison of the results of the $2D$ data at beginning of the process. Sub-figures (a) and (b) represent the realistic simulation of the contour plot. Then, sub-figures (c) and (d) are the predicted values with less than 20% of the training set. Finally, sub-figures (e) and (f) show the best result obtained with more than 70% of the training set.

However, if we consider the test data higher than 20% then we can observe that the model performs significantly better (see sub-figures 4.13 (c, and d). In this case, the prediction values are close to the values of the simulated data. This refers to the comparison between the two figures with the same time steps 4.13 (a, and c) and 4.13 (b, and f). The results show in all cases a globally good qualitative agreement with the results carried out by the FVM simulation approach. Overall, the results obtained show that the PINN approach can accurately predict or estimate the velocity fields and the unknown parameters of the PDE. Furthermore, this supports the effectiveness of the method such that FCNN informed by the NSE can be a valid candidate to substitute the numerical solution obtained by the FVM method which simulates the FSW process.

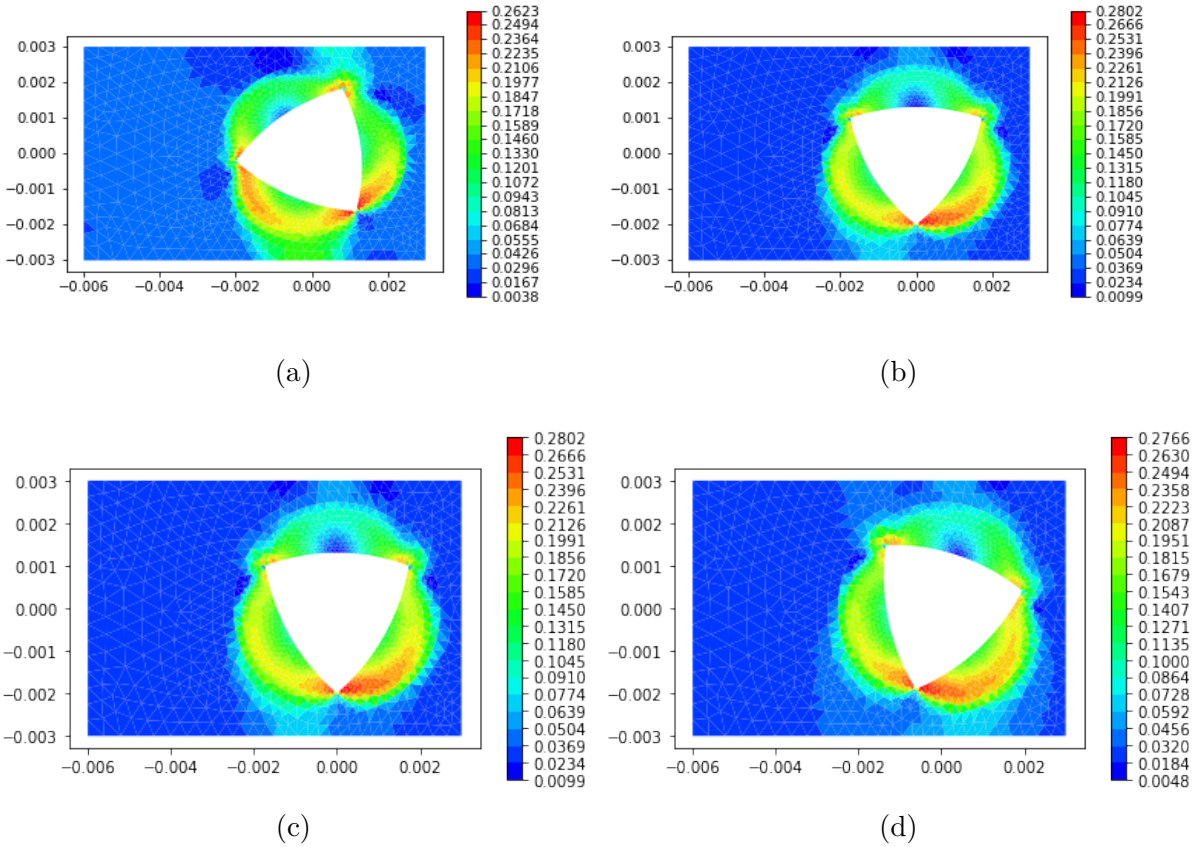


Figure 4.14: A comparison of the results of the $2D$ data at the middle and end of the process. The first two contours plots (a) and (b) represent the data realistic simulation and the last two contours plot (c) and (d) represent the best result obtained

4.6.3 Discussions

In this study, we focused on the resolution of PDEs by the PINN model using the data from numerical simulation methods this is a supervised learning task. In fact, the objective is to perform a solution that can be replaced a numerical model. Given the characteristics of the Navier-Stokes Equation, a PINN is formed and trained on the simulated data. Furthermore, we investigate the performance of the model through the RMSE, and the following discussion is presented.

- PINN is the network that provides the ability to encode model equations, such as PDE, as part of the FCNN itself.
- The principal aim of the PINN model is to integrate an adequate regularization of the physical knowledge into the FCNN. To improve the learning of the solutions of the PDEs, we inject the physical constraints in the network, which guides the model to the optimal solutions from limited information. The regularization of the physical information is performed through the automatic differentiation of the Pytorch framework. This is approximated because the NSE does not have a known solution.
- The regularization has an important role in the PINN, besides, the regularization related to the physics constraints we have also developed the dropout regularization. In future works, it will be interesting to explore other methods of regularization such as the L_1 or L_2 norm. In particular, the L_1 norm approach is robust to the interference of anomalous or biased data. These approaches improve the knowledge of the methods related to PINNs.
- For the purpose of PINNs training, the simulated $2D$ data are randomly selected in the space-time domain. In addition, traditional numerical methods such as FVMs require the discretization of the PDEs while PINNs can directly learn the solutions of these equations from a small amount of data without the necessity of this discretization. Moreover, PINNs can be developed and adapted to solve other problems that consider the discretization of the equations (i.e., numerical weather prediction models). Consequently, this model favors the implementation of several case studies, especially the development of computational fluid dynamics and scientific computing.
- The proposed framework has several advantages, such as the fact that it does not consider the discretization of the PDEs, where the solutions are computed by automatic differentiation. Moreover, the introduction of the regularization process allows to guide or constrain the model to follow the phenomenology of the system, which reduces the risks of overfitting, thus improving the generalization performances and guaranteeing solutions that are consistent with the considered systems. However, PINNs are confronted with various difficulties, such as the resolution of PDEs via FCNNs being highly dependent on the training data, which often requires a long training time when the quality of the data is poor. In addition, the performance of the model can also depend on the hyperparameters of the model. To address these issues, it is important to use

high-quality data to reduce training time and to exploit methods for the automatic selection of regularization parameters.

- PINNs allow the choice of network architecture, automatic differentiation algorithm, and advanced machine learning software with parallel processing capabilities by CPU and GPU.

- In order to improve the PINNs model, we will study the behavior and the weight of the pressure (p) during the training of the model. Despite a logarithmic transformation of the loss function related to this variable, we found that the problem is not completely solved. This has a significant impact on the combined loss function. Furthermore, in future works, it is important to include the boundary conditions in the PINN model. This factor contributes more knowledge about the model, especially about the conditions related to the tool.

- We study 2D-dimensional PDE problems, but in future works, the method should be applied to 3-dimensional simulation problems where the spatial coordinates are in the x-axis, y-axis, and z-axis

4.7 Conclusion

In this chapter, we propose an efficient and robust computational framework named PINN. The developed neural network addresses a solid mechanics problem where the viscosity is expressed by the Navier-Stokes equation which depends on the strain rate. In addition, the neural network aims to learn the process in the transient regime and it is able to predict the whole process from the beginning to the end of the regime (stationary). Our approach is composed of two major modules, the first one is based on the numerical simulation method known as the FVM. This module is based on the simulation of the realistic data of the FSW process with respect to the conditions of the tool parameters. The second module called PINN is composed of several blocks in particular the FCNNs, the combined loss function based on the physical model automatic differentiation, and the feedback mechanism. To inform the model, the physical laws or constraints included in the PDEs are introduced to the FCNNs in the form of regularization. In addition, dropout regularization is used to reduce or attenuate the noise factor related to the data.

Compared to the traditional NNs, this method allows the model to be trained from a small number of observations. Moreover, by informing the model with the NSE, the solutions of the PDEs can be learned better. This effectively reduces the search field of the algorithm and guides it to an optimal solution. The PINN is considered a new method of solving PDEs by NNs that combines physical information and data where the PDE solutions are calculated by automatic differentiation. In addition, the results obtained are robust and generalizable due to the ability to approximate the functions of the NNs and the NSE. To evaluate the performance of the model when predicting the velocities and pressure of material deformation during the welding process, we used the RMSE which is based on the combined

loss function. This metric allows us to find the optimal network for the estimation of the hyper-parameters in the fine-tuning phase. According to the obtained experimental results and the state-of-the-art developed on PINNs, we expect that PINNs can substitute numerical simulation solutions that require considerable computational time and memory. Moreover, they will considerably impact the study of PDEs resolution and thus support the development of scientific computing.

However, there are still many challenges and improvements related to PINNs, including, the questions related to the PINNS theories which remain to be solved. In particular, we will focus on the introduction of boundary conditions in NNs and we will perform a comparative analysis (accuracy and computational time) of the performance differences between PINN-based methods and FVM methods.

Chapter 5

General Conclusions and Futures works

Sommaire

5.1 Summary of Contributions 136

5.1.1 Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis 136

5.1.2 Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP Explainable 137

5.1.3 Physical-Informed Neural Networks and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding 138

5.2 Limitations of the Proposed Approaches 139

5.3 Suggestions for Future Works 140

In this last chapter, we will briefly present the summaries, suggestions, limitations, and future work concerning the development of sophisticated decision-support tools. Prior to any development, we thought it appropriate to present the different industrial revolutions, their main characteristics, and specificities. It emerges that Industry 4.0 contributes significantly to the new challenges in the organization of customized production resources. Thus, it is identified as a new way of planning and organizing all resources of the factory. This is environmentally responsible and saves energy and raw materials. We have noted that manufacturers are confronted with several challenges in an increasingly competitive market. To meet these new challenges, an approach consists of digitalizing or automating the industry with the introduction of cyber-physical systems, communicating sensors, and intelligent and autonomous robots. Beyond these challenges, we have presented the nine pillars or technologies that constitute the foundation of the Factory of the Future.

In particular, we focused on a combination of these technologies to highlight the approaches or strategies of industrial maintenance. We mention that corrective and preventive maintenance does not fully address the issues related to I4.0. This is because scheduled maintenance or corrective maintenance does not ensure the continuing operation of the equipment. Consequently, we have focused on the predictive maintenance strategy which consists of anticipating anomalies, possible machine malfunctions, or computing the life duration of a component. This strategy can be used as a recommended or feedback system. We have also shown the main steps for the implementation of this strategy of maintenance.

To predict the failures of a system several approaches can be exploited. We have focused on data-driven and hybrid approaches. These frameworks have an important function in resolving the tasks in PdM 4.0 activities. In particular, PdM 4.0 and AI-based systems offer new opportunities for optimizing production processes and reducing maintenance costs. These maintenance costs are very often due to equipment that suffers from deterioration without being identified in advance. A sudden shutdown of the machines can also lead to huge losses in the production system.

To address the objectives and research questions of the study, several methods and frameworks were developed, including data mining and Deep Learning techniques. In particular, we have developed decision support tools based on deep learning models. We explored DNN models, hybrid models combined with explanation approaches, and models informed by physical laws.

5.1 Summary of Contributions

5.1.1 Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis

Firstly, we have performed a state-of-the-art study about AI techniques applied to PdM in I4.0. To address the main questions related to industrial-based systems, we exploited the data mining technique named Bibliometrics. The bibliometric analysis was achieved using the processing and analysis tools Biblioshiny, and data visualization tools such as VOSviewer and Power BI. We collected 4096 scientific documents published between 2000 and 2021 in the scientific database WoS. This bibliometric study provides an overview of the most important concepts, topics, and application areas of AI. In addition, the analyses highlight the AI methods, their particularities, the main trends, results, or performances. Furthermore, the descriptive analyses focus on the publication trends, sources, articles, authors, co-authors, references, affiliations, and most productive countries. In addition, we have visualized networks or clusters of collaboration between authors and institutions. We have highlighted the ethical, trusting, transparent, and socio-economic impacts of using these models. We also presented the potential, the main challenges, and the future research directions of AI systems.

The most important results show an exponential use of AI techniques in Industry 4.0, especially in PdM and anomaly detection. We observe that the gap between developed and third countries is increasing in terms of research and industrialization. However, we showed that the emerging themes in AI are DL, and ML techniques. In addition, the most commonly used models in surveillance, diagnosis, prognosis, anomaly detection, data denoising, pattern recognition or signal processing, transfer learning or random fake data generations are mainly CNN, DNN, ANN, LSTM, AE, and GAN. Their models can be used for modeling, identification, optimization, prediction, and control of complex systems. Despite their performances and their numerous applications, the models have limitations in practical cases. For example, the real-time prediction, and the dependence of some models on the quality of the data.

Problems of instability and over-fitting in the presence of missing or biased data models or in the presence of complex, unbalanced, or mislabeled classes. Models such as CNNs and DNNs have complex architectures, which can require significant GPU and computational resources to estimate the parameters. In this context, real-time or online analyses can become very complicated due to the high computation time and complexity of these models.

These models are known as "black box" models, which means that the internal structure and decision rules are unknown or not explicable. This lack of explicability can be a real problem for the generalization of these models in the industry. The use of AI technologies could face some challenges such as operation, organization, adaptability, machine-machine interactions, human-machine interactions, cybersecurity (risky attacks), online analysis, real-time data collection, and data quality. New trends in AI approaches have the following: development of hybrid multimodal models, visual reasoning, XAI explainability, and feature selection for real-time predictive maintenance in the industry.

5.1.2 Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP Explainable

Following a detailed analysis of traditional AI models applied to PdM. We have developed a decision support tool to improve maintenance operations or activities. The resulting tool is a hybrid framework that addresses the condition monitoring (CM) tasks of a hydraulic system based on multi-sensor data. In addition, this tool aims to monitor and predict the different component conditions (coolers, valves, internal pump leaks, or the state of the hydraulic accumulator).

Prior to any development, we have provided answers to the problem of data quality the proposed solutions are the following; Manual data labeling, artificial resampling data, data augmentation by using DL, clustering technique, simulation of "fake" or realistic data, and generation of data or features by using transfer learning (TL) technique. Furthermore, we presented the benefits and strategies of maintenance such as manufacturing process optimization, diagnosis, prognosis, and monitoring of interconnected machines. In particular, we have shown the different tasks that can be covered by CM techniques (monitoring of machine components, sensors, and the CM framework itself). To perform the prediction or monitoring tasks, we have studied the different approaches. For our case study, we focused on the data-driven approach. The model used is a DNN which can be applied to several tasks (fault classification, diagnosis, and non-linear problem solving). Furthermore, the model is robust to learning unbalanced multi-class classification tasks and noisy data.

We have developed a hybrid framework consisting of two main modules: a multi-class DNN classification model with unbalanced classes combined with a DeepSHAP explainable model. We have shown that the classification model can be directly applied to the data in order to capture or extract deep features. In this case, a feature selection operation is not necessary. We evaluated the performance of the model via Accuracy, F1-Score, Precision, Recall, confusion matrices, and classification errors. The model training performs well in

the presence of unbalanced data because we regularized the loss function and we applied the cross-validation technique. In particular, the results show that the classification rate for each target variable or component of the hydraulic system is close to 100% including cooler conditions (99.87%), valve conditions (99.60%), pump internal leakage (99.09%), and stable state (94.17%), and hydraulic accumulator conditions (88.60%).

Although DNNs are powerful, their reliability as "black box" models remains problematic because we are not able to explain the decision rules. To address this explainability paradigm, we focused on deep learning reasoning using the XAI explanation approach called DeepSHAP. This approach provides an explanation of the DNN results. In addition, DeepSHAP discriminates the model outputs more efficiently by providing both local and global explanations. The extracted knowledge provides additional information on the local and global importance or contribution of each variable in the DNN model decision process. This knowledge is then used to optimize the prediction model. The DeepSHap results show that global monitoring of the sensors is not necessary. As an example, for the monitoring of the internal pump and hydraulic accumulator we may only monitor respectively seven sensors (P1, CE1, TS4, TS1, TS2, TS3, and SE) and seven sensors (P1, TS4, SE1, TS1, CE1, TS3, and TS2). These results show that the DeepSHAP approach can perform the role of feature selection in the context of machine learning. In this context, we are monitoring a reduced number of sensors (6 instead of 17), which probably reduces the cost of monitoring industrial systems.

Additionally, explainability can also suggest possible indicators to be monitored. As an example, the cooling condition of the hydraulic system is most likely conditioned by the quantity of cooling pumped, the pressure, the engine power, and the temperature of the cooler to maintain the pump at a normal temperature. The explanation framework helps to improve the decision-makers understanding and interpretation of the IA algorithms and their decision-making rule. This can support implementation and confidence in these models in maintenance applications.

5.1.3 Physical-Informed Neural Networks and Numerical Simulation of Thermomechanical Process: Application to the Friction Stir Welding

In the preceding sub-section, we highlighted the explanatory methods for extracting knowledge related to the DNN model prediction rules. In this conclusion, we will address another important topic of this thesis, which is to introduce the knowledge of the process during the training of the neural network. We have shown that AI models can be applied to the FSW process. We have developed a hybrid model called Physics Informed Neural Network (PINN) that is applied to simulated data. This framework addresses a solid mechanics problem where the viscosity in the Navier-Stokes equation is expressed by the Norton-Hoff law which depends on the strain rate. The resulting framework is composed of two modules: Numerical simulation and PINN model. We have reviewed the state-of-the-art numerical simulation methods and have focused on the Finite Volume Method (FVM). This approach consists in determining the local fields to be assigned to each element so that the global field obtained by juxtaposing these local fields is close to the solution to the initial problem. Although the

considered fields may be distorted, these fields are close to the solution to the initial problem. The realistic data used to train the PINN model is therefore obtained from this numerical simulation. Considering that FSW is a computationally intensive process, we proposed a second module that consists of a fully connected neural network (FCNN) informed by the physical law. By informing the model with NSE, the solutions of the DPEs can be better trained. This effectively reduces the search space of the algorithm and guides it toward an optimal solution. The PINN is therefore considered a new method for solving PDEs and the solutions are computed by automatic differentiation. The model obtained helps to learn the transient phase of the process.

By including knowledge that is physically constrained or regularisation terms, when training the model we obtain a compound loss function. In addition, dropout regularisation is used to reduce or attenuate the noise factor associated with the data. To evaluate the performance of the network model in predicting material strain rates, velocities, and pressure during the welding process, we used the RMSE. This metric helps to find the optimal network for the estimation of hyperparameters in the fine-tuning phase. Based on the experimental results obtained, we consider that PINNs can replace numerical simulation solutions that require considerable computation time and memory. Thus, Finally, modeling with PIN models should yield significant time saving due to its memory effect.

5.2 Limitations of the Proposed Approaches

Regarding the bibliometric analysis, we have developed our query using keywords related to the contexts of the study. However, we cannot guarantee that all the documents collected take into account all the research fields of IA applied to PdM. This observation could be negligible if we had considered multiple databases or writing languages. Furthermore, combining bibliometric approaches might be necessary to improve the results. Furthermore, we mentioned that the maintenance framework can cover several tasks. However, the DNN classifier that we have developed does not take into account possible sensor failures, incomplete data, and multi-modal data (images, texts, etc..). In addition, we have not considered the physical aspects of the system (the case of the hydraulic system). However, the explanation approach provides additional information that contributes to a better understanding of the processes.

In addition, there are still many challenges and improvements related to PINNs, including questions about the PINNS theories that remain to be resolved. Moreover, when training the model we did not take into account the initial and boundary conditions. In summary, we have proposed two hybrid frameworks. The first framework extracts knowledge from the prediction model. The second one informs the prediction model with the understanding of the process. This knowledge facilitates the training of the models, thus, the obtained models are performing robustly and are generalizable.

5.3 Suggestions for Future Works

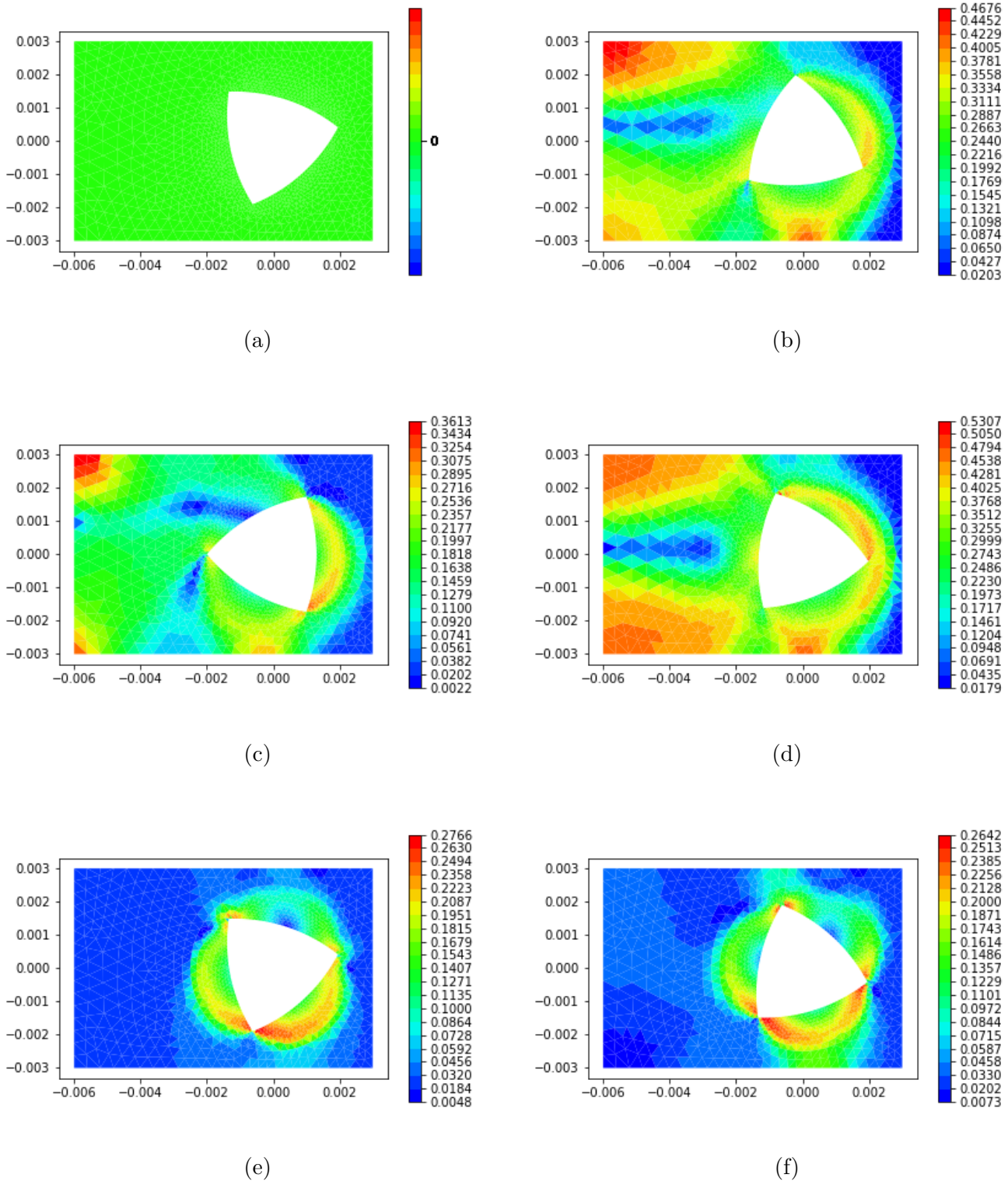
Concerning the bibliometric analysis, we can improve the results by refining the initial query or by adding other keywords to the query. To collect more documents, we will also exploit several scientific databases (Scopus, Springer, Google scholar, Science Direct, and IEEE). To further improve our analysis, we will also take into account documents (articles, thesis, books, etc.) written in languages other than English. Finally, we will also exploit a combination of several traditional bibliometric methods to improve the accuracy of the results.

Regarding condition monitoring, we will study the robustness of the DNN to missing or mislabelled data. In addition, we will integrate information related to the lifetime of the sensors in the framework by considering the abnormal operation (aging) of the sensor. In addition, we will explore and integrate data related to air, oil, and water pollution into the classification model. These variables could contain information related to the failure of hydraulic systems. Finally, we will be able to combine several explanatory AI approaches. Concerning the applicability, we will validate or confirm the obtained results by industrial experts.

To improve the PINNs model, we will study the behavior and weight of the pressure during the training of the model and propose further transformations of the loss function related to this variable. In addition, we will include the initial and boundary conditions when training the PINNs. These factors will provide more knowledge about the model, in particular regarding the tool conditions. We will evaluate the training of PINNs on three-dimensional data where the spatial coordinates are in the x-axis, y-axis, and z-axis.

Appendix A

Appendix - Some results of PINN Module



(a) (b)

(c) (d)

(e) (f)

Figure A.1: Results of realistic data (simulation of the $2D$ FSW process). Contour plot represent the evolution of the velocities and pressure subsequently at the beginning. The time steps are respectively equal to 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec).

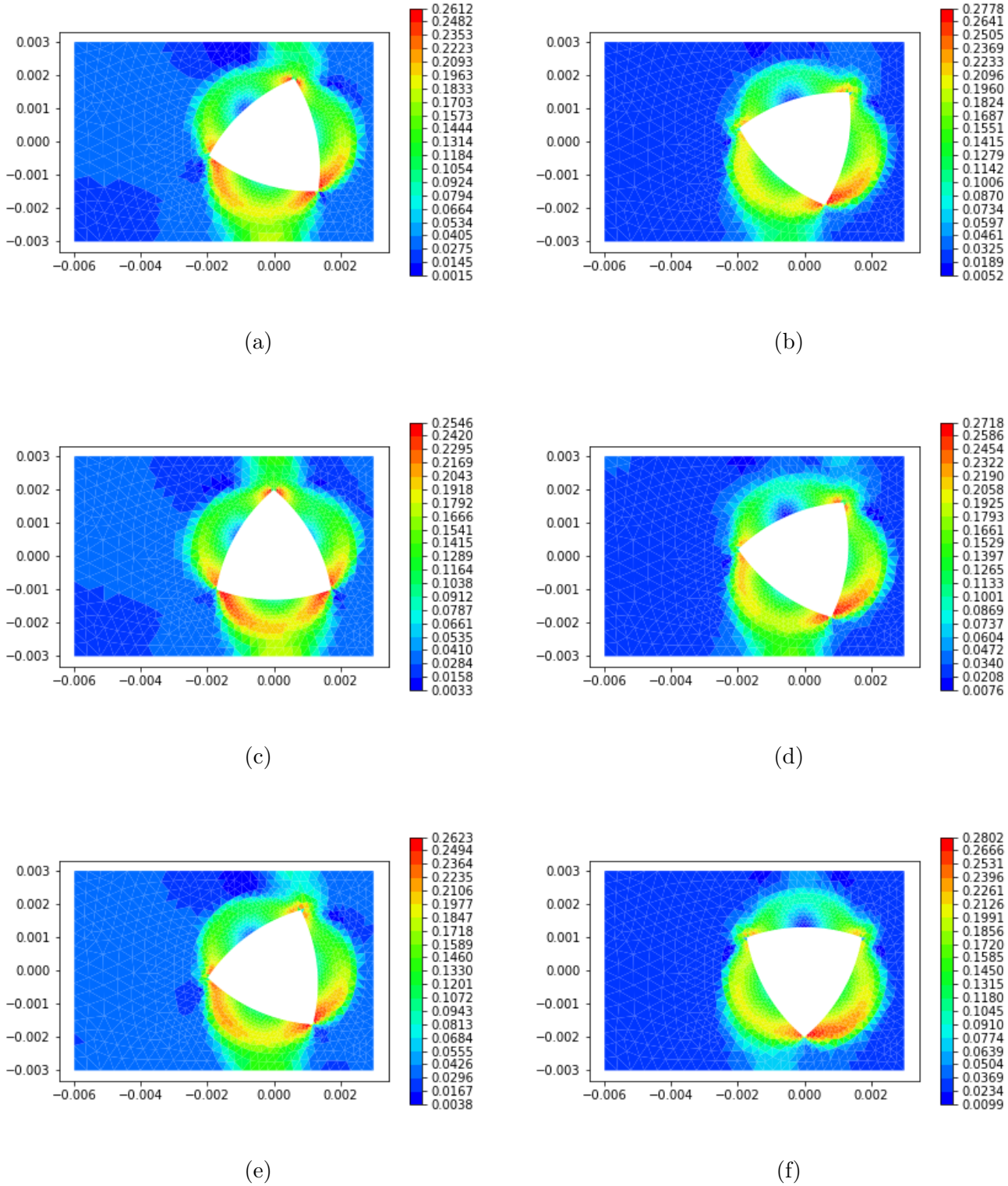
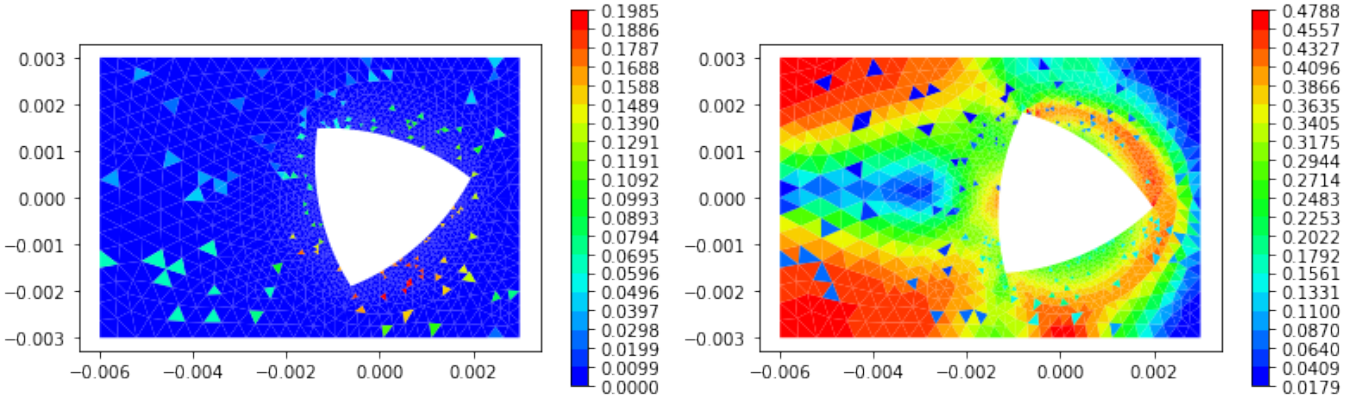
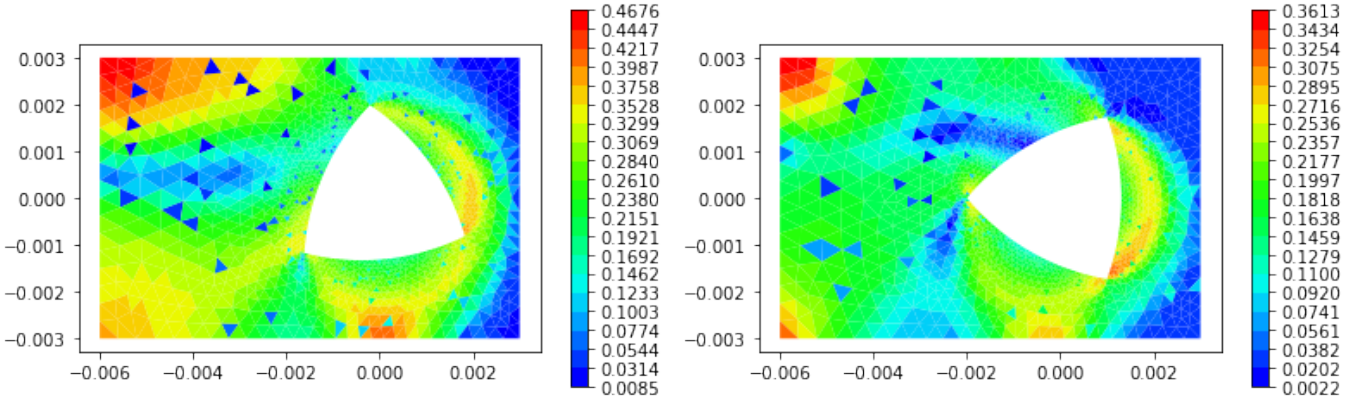


Figure A.2: Results of realistic data (simulation of the $2D$ FSW process). Contour plot represent the evolution of the velocities and pressure subsequently at the middle of the process. The time steps are respectively equal to 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec).



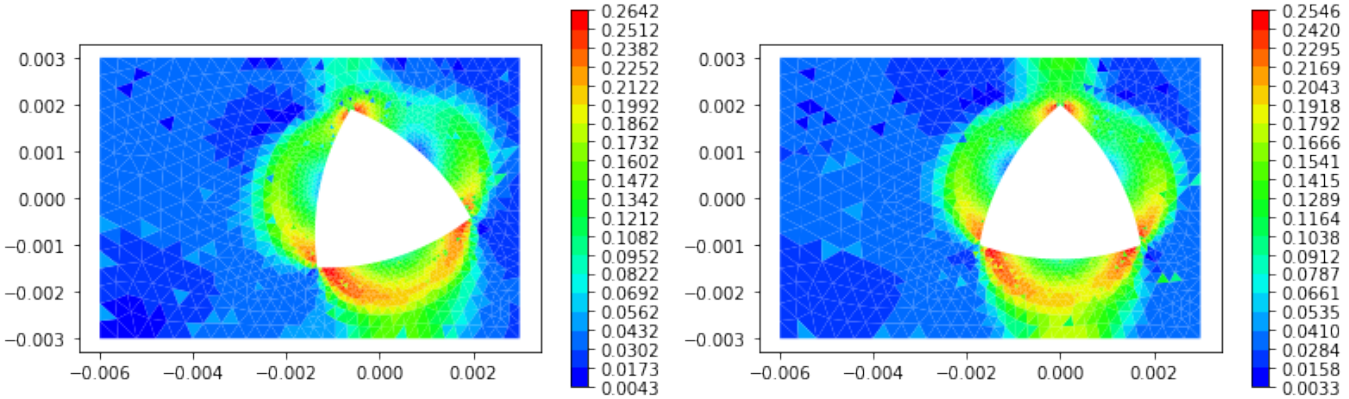
(a)

(b)



(c)

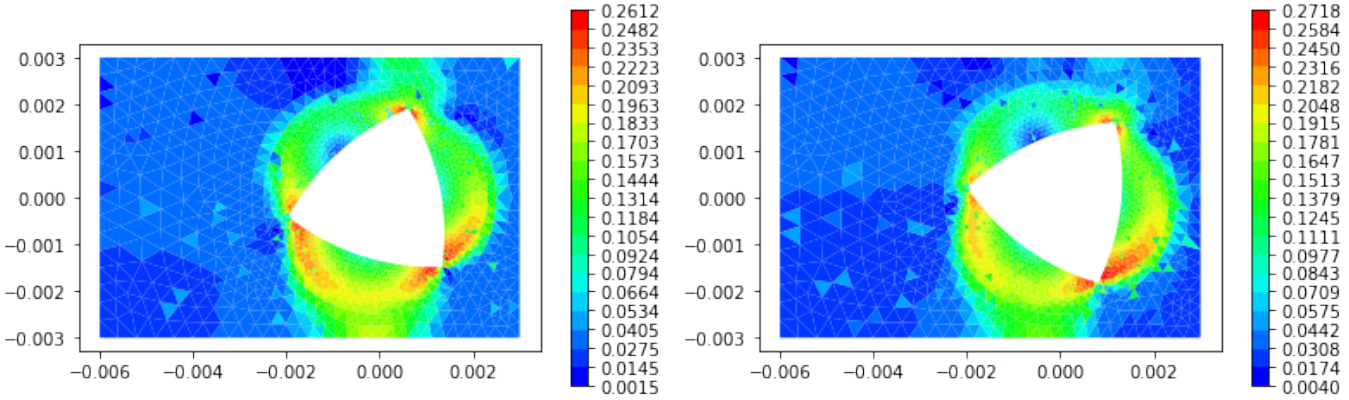
(d)



(e)

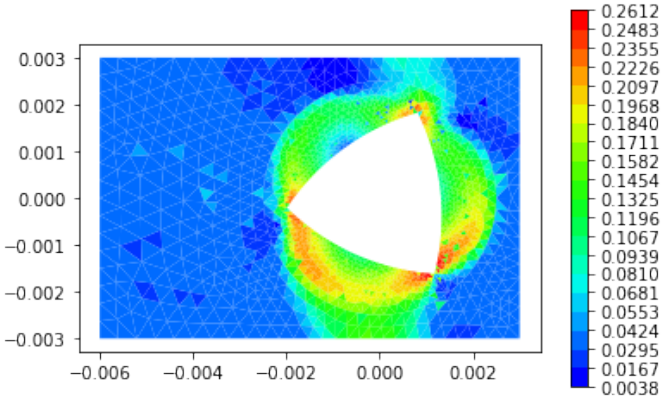
(f)

Figure A.3: Predicted values at the beginning of the process with less than 20% of the training set. The time steps are equal o 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec.

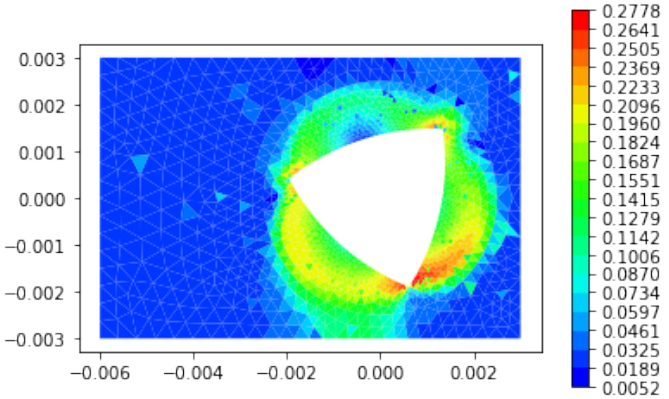


(a)

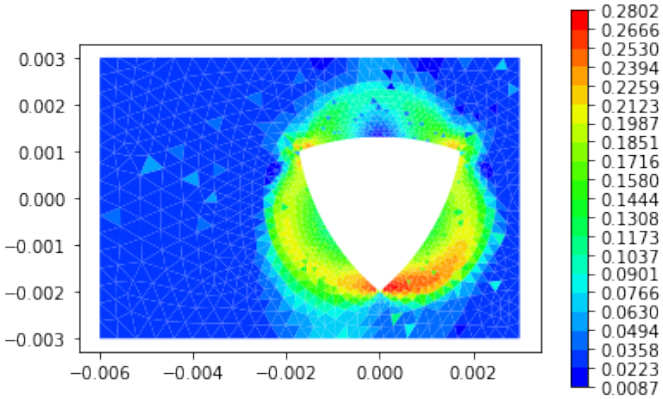
(b)



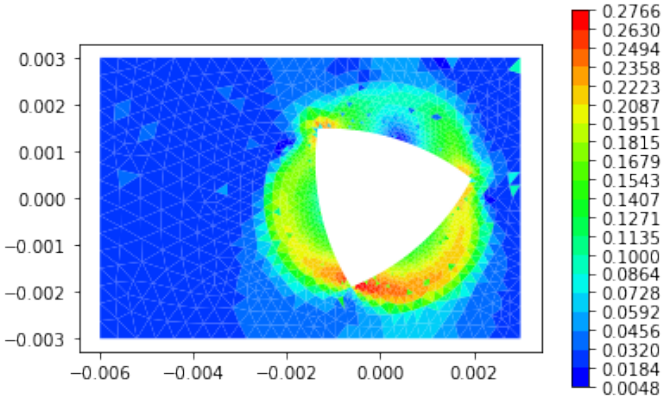
(c)



(d)



(e)



(f)

Figure A.4: Predicted values at the beginning of the process with less than 20% of the training set. The time steps are equal 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec).

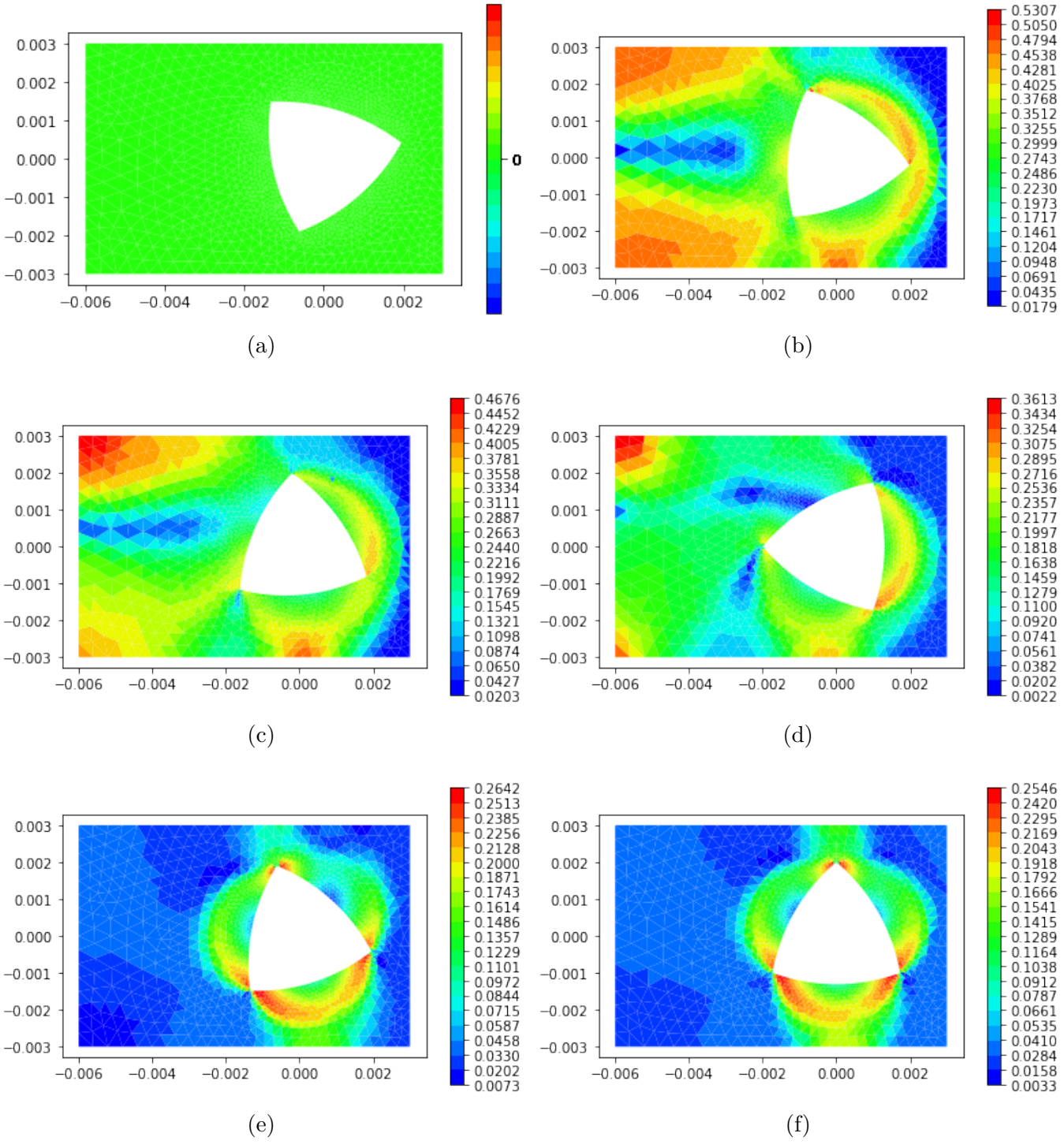
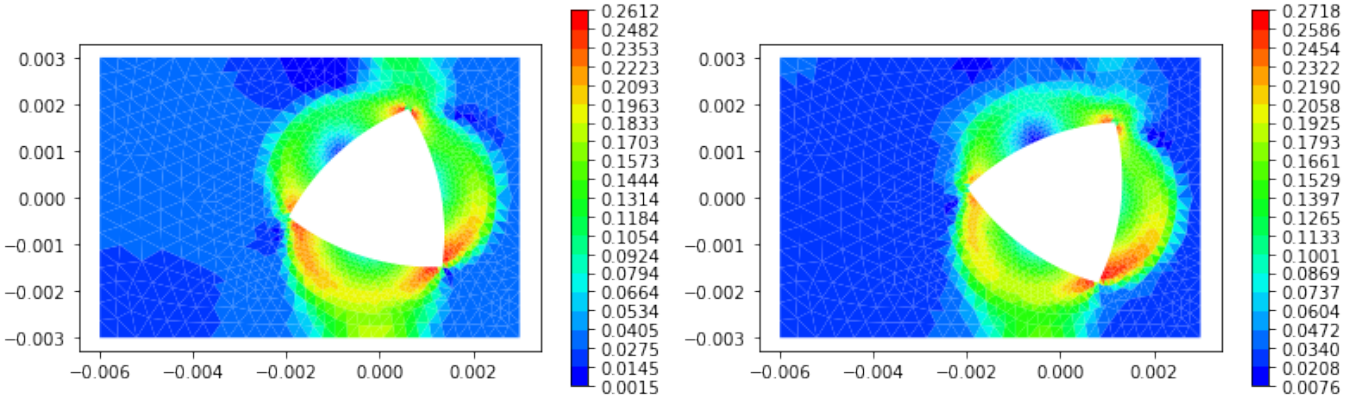
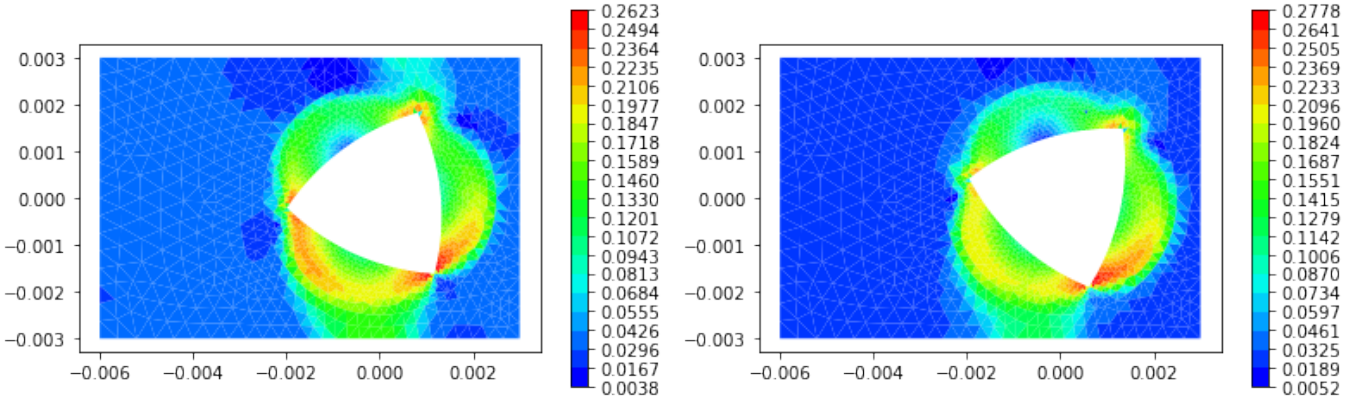


Figure A.5: The best result obtained with more than 70% of the training set. Predicted values at the beging of the process. The time steps are equal to 0.00, 0.0015, 0.0030, 0.0060, 0.0720, and 0.07350 sec



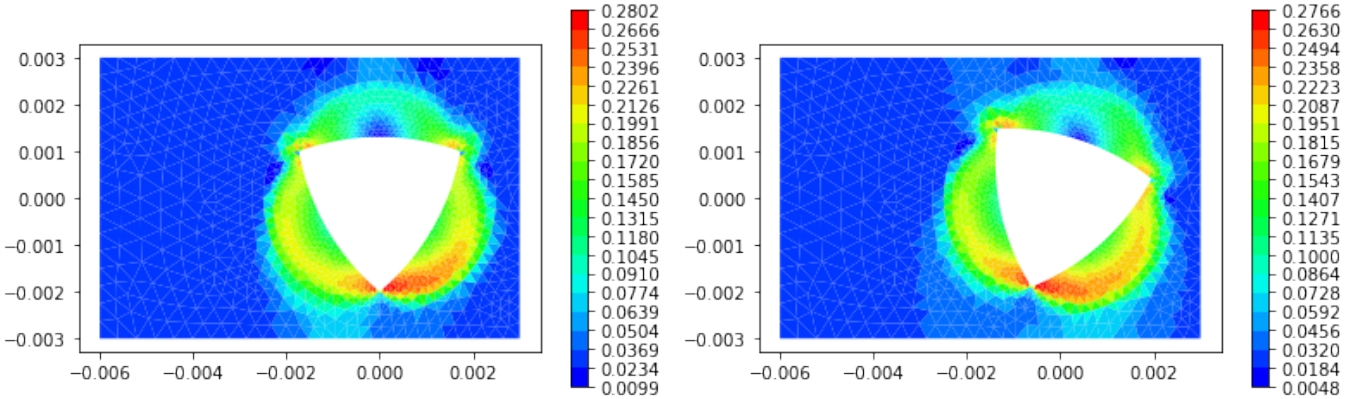
(a)

(b)



(c)

(d)



(e)

(f)

Figure A.6: The best result obtained with more than 70% of the training set. Predicted values at the end of the process. The time steps are equal to 0.0750, 0.0765, 0.14550, 0.1470, 0.14850, and 0.150 sec

Bibliography

[1] E. Oztemel and S. Gursev, "Literature review of industry 4.0 and related technologies," *Journal of Intelligent Manufacturing*, vol. 31, no. 1, pp. 127–182, 2020 (cit. on pp. 4, 19).

[2] Y. Lu, "Cyber physical system (cps)-based industry 4.0: A survey," *Journal of Industrial Integration and Management*, vol. 2, no. 03, p. 1 750 014, 2017 (cit. on pp. 4, 22).

[3] T. Bidet-Mayer, *L'industrie du futur: une compétition mondiale*. Presses des MINES, 2016, vol. 14 (cit. on p. 4).

[4] M. Rüßmann, M. Lorenz, P. Gerbert, *et al.*, "Industry 4.0: The future of productivity and growth in manufacturing industries," *Boston Consulting Group*, vol. 9, no. 1, pp. 54–89, 2015 (cit. on pp. 4, 7, 19).

[5] D. Palka and J. Ciukaj, "Prospects for development movement in the industry concept 4.0," *Multidisciplinary Aspects of Production Engineering*, vol. 2, 2019 (cit. on pp. 4, 6).

[6] R. Anderl, "Industrie 4.0–technological approaches, use cases, and implementation," *at-Automatisierungstechnik*, vol. 63, no. 10, pp. 753–765, 2015 (cit. on pp. 4, 19).

[7] F. F. Adedoyin, F. V. Bekun, O. M. Driha, and D. Balsalobre-Lorente, "The effects of air transportation, energy, ict and fdi on economic growth in the industry 4.0 era: Evidence from the united states," *Technological Forecasting and Social Change*, vol. 160, p. 120 297, 2020 (cit. on pp. 4, 19).

[8] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & information systems engineering*, vol. 6, no. 4, pp. 239–242, 2014 (cit. on pp. 4, 19).

[9] J. Wan, H. Cai, and K. Zhou, "Industrie 4.0: Enabling technologies," in *Proceedings of 2015 international conference on intelligent computing and internet of things*, IEEE, 2015, pp. 135–140 (cit. on pp. 4, 19).

[10] P. Radanliev, D. De Roure, J. R. Nurse, *et al.*, "New developments in cyber physical systems, the internet of things and the digital economy–discussion on future developments in the industrial internet of things and industry 4.0," 2019 (cit. on pp. 4, 19).

[11] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: State of the art and future trends," *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, 2018 (cit. on pp. 5, 19, 22, 59).

[12] E. Manavalan and K. Jayakrishna, "A review of internet of things (iot) embedded sustainable supply chain for industry 4.0 requirements," *Computers & Industrial Engineering*, vol. 127, pp. 925–953, 2019 (cit. on p. 7).

[13] G. Aceto, V. Persico, and A. Pescapé, "Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0," *Journal of Industrial Information Integration*, vol. 18, p. 100 129, 2020 (cit. on p. 7).

[14] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, p. 106 889, 2020 (cit. on pp. 9, 19).

[15] M. E. Porter and J. E. Heppelmann, "How smart, connected products are transforming competition," *Harvard business review*, vol. 92, no. 11, pp. 64–88, 2014 (cit. on pp. 9, 19).

[16] S. Sajid, A. Haleem, S. Bahl, M. Javaid, T. Goyal, and M. Mittal, "Data science applications for predictive maintenance and materials science in context to industry 4.0," *Materials Today: Proceedings*, 2021 (cit. on pp. 9, 19, 22).

[17] T. Welte, "A rule-based approach for establishing states in a markov process applied to maintenance modelling," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 223, no. 1, pp. 1–12, 2009 (cit. on pp. 9, 25, 59).

[18] A. Sergaki and K. Kalaitzakis, "A fuzzy knowledge based method for maintenance planning in a power system," *Reliability Engineering & System Safety*, vol. 77, no. 1, pp. 19–30, 2002 (cit. on pp. 9, 25, 59).

[19] R. Yu, B. Iung, and H. Panetto, "A multi-agents based e-maintenance system with case-based reasoning decision support," *Engineering applications of artificial intelligence*, vol. 16, no. 4, pp. 321–333, 2003 (cit. on pp. 9, 25, 59).

[20] T. A. N. Heirung and A. Mesbah, "Input design for active fault diagnosis," *Annual Reviews in Control*, vol. 47, pp. 35–50, 2019 (cit. on pp. 9, 26, 59).

[21] G. Zhou, W. Feng, Q. Zhao, and H. Zhao, "State tracking and fault diagnosis for dynamic systems using labeled uncertainty graph," *Sensors*, vol. 15, no. 11, pp. 28 031–28 051, 2015 (cit. on pp. 9, 26, 59).

[22] E. Benowitz, "The curiosity mars rover's fault protection engine," in *2014 IEEE International Conference on Space Mission Challenges for Information Technology*, IEEE, 2014, pp. 62–66 (cit. on pp. 9, 26, 59).

[23] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Computer Science*, vol. 60, pp. 708–713, 2015 (cit. on pp. 9, 26, 59).

[24] S. A. Taqvi, L. D. Tufa, H. Zabiri, S. Mahadzir, A. S. Maulud, and F. Uddin, "Artificial neural network for anomalies detection in distillation column," in *Asian simulation conference*, Springer, 2017, pp. 302–311 (cit. on pp. 9, 59).

[25] Y Dhanalakshmi and I. R. Babu, "Intrusion detection using data mining along fuzzy logic and genetic algorithms," *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 27–32, 2008 (cit. on pp. 9, 26, 59).

[26] F.-K. Wang and T. Mamo, "Hybrid approach for remaining useful life prediction of ball bearings," *Quality and Reliability Engineering International*, vol. 35, no. 7, pp. 2494–2505, 2019 (cit. on p. 9).

[27] H. Ye, F. Cao, and D. Wang, "A hybrid regularization approach for random vector functional-link networks," *Expert Systems with Applications*, vol. 140, p. 112 912, 2020 (cit. on p. 9).

[28] W. Luo, T. Hu, Y. Ye, C. Zhang, and Y. Wei, "A hybrid predictive maintenance approach for cnc machine tool driven by digital twin," *Robotics and Computer-Integrated Manufacturing*, vol. 65, p. 101 974, 2020 (cit. on pp. 9, 25, 109).

[29] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," in *2016 49th Hawaii international conference on system sciences (HICSS)*, IEEE, 2016, pp. 3928–3937 (cit. on p. 19).

[30] F. Orellana and R. Torres, "From legacy-based factories to smart factories level 2 according to the industry 4.0," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 4-5, pp. 441–451, 2019 (cit. on p. 19).

[31] M. Brettel, M. Klein, and N. Friederichsen, "The relevance of manufacturing flexibility in the context of industrie 4.0," *Procedia Cirp*, vol. 41, pp. 105–110, 2016 (cit. on p. 19).

[32] O. Szymańska, M. Adamczak, and P. Cyplik, "Logistics 4.0-a new paradigm or set of known solutions?" *Research in Logistics & Production*, vol. 7, 2017 (cit. on p. 19).

[33] J. Fischer, B. Obst, and B. Lee, "Integrating material flow simulation tools in a service-oriented industrial context," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, IEEE, 2017, pp. 1135–1140 (cit. on p. 19).

[34] M. Sony and S. Naik, "Critical factors for the successful implementation of industry 4.0: A review and future research direction," *Production Planning & Control*, vol. 31, no. 10, pp. 799–815, 2020 (cit. on p. 19).

[35] D. Lund, C. MacGillivray, V. Turner, and M. Morales, "Worldwide and regional internet of things (iot) 2014–2020 forecast: A virtuous circle of proven value and demand," *International Data Corporation (IDC), Tech. Rep*, vol. 1, p. 9, 2014 (cit. on p. 19).

[36] E. Sezer, D. Romero, F. Guedea, M. Macchi, and C. Emmanouilidis, "An industry 4.0-enabled low cost predictive maintenance approach for smes," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, 2018, pp. 1–8 (cit. on p. 19).

[37] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring," *Computer Networks*, vol. 101, pp. 192–202, 2016 (cit. on pp. 19, 59).

[38] D. Trotta and P. Garengo, "Industry 4.0 key research topics: A bibliometric review," in *2018 7th international conference on industrial technology and management (ICITM)*, IEEE, 2018, pp. 113–117 (cit. on p. 20).

[39] M. L. H. Souza, C. A. da Costa, G. de Oliveira Ramos, and R. da Rosa Righi, "A survey on decision-making based on system reliability in the context of industry 4.0," *Journal of Manufacturing Systems*, vol. 56, pp. 133–156, 2020 (cit. on p. 20).

[40] P. K. Muhuri, A. K. Shukla, and A. Abraham, "Industry 4.0: A bibliometric analysis and detailed overview," *Engineering applications of artificial intelligence*, vol. 78, pp. 218–235, 2019 (cit. on pp. 20, 30).

[41] M. Mariani and M. Borghi, "Industry 4.0: A bibliometric review of its managerial intellectual structure and potential evolution in the service industries," *Technological Forecasting and Social Change*, vol. 149, p. 119 752, 2019 (cit. on p. 20).

[42] C. Cézanne, E. Lorenz, and L. Saglietto, "Exploring the economic and social impacts of industry 4.0," *Revue d'économie industrielle*, no. 1, pp. 11–35, 2020 (cit. on p. 20).

[43] K. Ejsmont, B. Gladysz, and A. Kluczek, "Impact of industry 4.0 on sustainability—bibliometric literature review," *Sustainability*, vol. 12, no. 14, p. 5650, 2020 (cit. on p. 20).

[44] V. GRUBISIC, J. AGUIAR, and Z SIMEU-ABAZI, "A review on intelligent predictive maintenance: Bibliometric analysis and new research directions," in *2020 International Conference on Control, Automation and Diagnosis (ICCAD)*, IEEE, 2020, pp. 1–6 (cit. on p. 20).

[45] M. A. Noman, E. S. A. Nasr, A. Al-Shayea, and H. Kaid, "Overview of predictive condition based maintenance research using bibliometric indicators," *Journal of King Saud University-Engineering Sciences*, vol. 31, no. 4, pp. 355–367, 2019 (cit. on p. 20).

[46] D. Yu, Z. Xu, and H. Fujita, "Bibliometric analysis on the evolution of applied intelligence," *Applied Intelligence*, vol. 49, no. 2, pp. 449–462, 2019 (cit. on p. 20).

[47] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, "Machine learning for industrial applications: A comprehensive literature review," *Expert Systems with Applications*, p. 114 820, 2021 (cit. on p. 20).

[48] P. Kamat and R. Sugandhi, "Bibliometric analysis of bearing fault detection using artificial intelligence," *Library Philosophy and Practice*, pp. 1–21, 2020 (cit. on p. 20).

[49] M. Aria and C. Cuccurullo, "Bibliometrix: An r-tool for comprehensive science mapping analysis," *Journal of informetrics*, vol. 11, no. 4, pp. 959–975, 2017 (cit. on pp. 20, 32, 33, 46).

[50] N. J. Van Eck and L. Waltman, "Software survey: Vosviewer, a computer program for bibliometric mapping," *scientometrics*, vol. 84, no. 2, pp. 523–538, 2010 (cit. on p. 20).

[51] A. Haleem and M. Javaid, "Additive manufacturing applications in industry 4.0: A review," *Journal of Industrial Integration and Management*, vol. 4, no. 04, p. 1 930 001, 2019 (cit. on p. 21).

[52] M. Javaid and A. Haleem, "Impact of industry 4.0 to create advancements in orthopaedics," *Journal of Clinical Orthopaedics and Trauma*, vol. 11, S491–S499, 2020 (cit. on p. 21).

[53] R. Strange and A. Zucchella, "Industry 4.0, global value chains and international business," *Multinational Business Review*, 2017 (cit. on p. 21).

[54] C. Stenström, A. Parida, U. Kumar, and D. Galar, "Performance indicators and terminology for value driven maintenance," *Journal of Quality in Maintenance Engineering*, 2013 (cit. on p. 22).

[55] K.-S. Wang, Z. Li, J. Braaten, and Q. Yu, "Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using anns," *Advances in Manufacturing*, vol. 3, no. 2, pp. 97–104, 2015 (cit. on pp. 22, 24).

[56] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 2518–2525 (cit. on p. 22).

[57] S. Ayad, L. S. Terrissa, and N. Zerhouni, "An iot approach for a smart maintenance," in *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, IEEE, 2018, pp. 210–214 (cit. on p. 22).

[58] M Haarman, P de Klerk, P Decaigny, *et al.*, "Predictive maintenance 4.0-beyond the hype: Pdm 4.0 delivers results," *PricewaterhouseCoopers and Mannovation*, 2018 (cit. on p. 22).

[59] E. J. De Visser, R. Pak, and T. H. Shaw, "From 'automation'to 'autonomy': The importance of trust repair in human–machine interaction," *Ergonomics*, vol. 61, no. 10, pp. 1409–1427, 2018 (cit. on p. 22).

[60] T Mcconnell, *Moral dilemmas [online]. usa: Stanford university: Center for the study of language and information*, 2014 (cit. on p. 24).

[61] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang, "Building ethics into artificial intelligence," *arXiv preprint arXiv:1812.02953*, 2018 (cit. on p. 24).

[62] N. Cointe, G. Bonnet, and O. Boissier, "Ethical judgment of agents' behaviors in multiagent systems," in *Proceedings of the 2016 international conference on autonomous agents & multiagent systems*, 2016, pp. 1106–1114 (cit. on p. 24).

[63] T. Saßmannshausen, P. Burggräf, J. Wagner, M. Hassenzahl, T. Heupel, and F. Steinberg, "Trust in artificial intelligence within production management–an exploration of antecedents," *Ergonomics*, pp. 1–18, 2021 (cit. on p. 24).

[64] C. Toro, C. Sanín, J. Vaquero, J. Posada, and E. Szczerbicki, "Knowledge based industrial maintenance using portable devices and augmented reality," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Springer, 2007, pp. 295–302 (cit. on p. 25).

[65] S. H. L. Mirhosseyni and P. Webb, "A hybrid fuzzy knowledge-based expert system and genetic algorithm for efficient selection and assignment of material handling equipment," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11 875–11 887, 2009 (cit. on p. 25).

[66] R. J. Sárfi, A. Solo, and D. Wilmington, "The application of fuzzy logic in a hybrid fuzzy knowledge-based system for multiobjective optimization of power distribution system operations," in *Proceedings of the 2005 International Conference on Information and Knowledge Engineering*, 2005, pp. 3–9 (cit. on p. 25).

[67] T. Ruppert and J. Abonyi, "Software sensor for activity-time monitoring and fault detection in production lines," *Sensors*, vol. 18, no. 7, p. 2346, 2018 (cit. on pp. 25, 107).

[68] M. Käßmeyer, R. Berndt, P. Bazan, and R. German, "Product line fault tree analysis by means of multi-valued decision diagrams," in *International GI/ITG Conference on Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*, Springer, 2016, pp. 122–136 (cit. on pp. 25, 107).

[69] T. Tinga, "Application of physical failure models to enable usage and load based maintenance," *Reliability Engineering & System Safety*, vol. 95, no. 10, pp. 1061–1075, 2010 (cit. on pp. 25, 108).

[70] H. Qin, Z. Xu, P. Li, and S. Yu, "A physical model approach to nonlinear vertical accelerations and mooring loads of an offshore aquaculture cage induced by wave-structure interactions," *Ocean Engineering*, vol. 197, p. 106 904, 2020 (cit. on pp. 25, 108).

[71] J. L. O. Coscia, M. Crasso, C. Mateos, A. Zunino, and S. Misra, "Predicting web service maintainability via object-oriented metrics: A statistics-based approach," in *International Conference on Computational Science and Its Applications*, Springer, 2012, pp. 29–39 (cit. on p. 25).

[72] S. Namuduri, B. N. Narayanan, V. S. P. Davuluru, L. Burton, and S. Bhansali, "Deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors," *Journal of The Electrochemical Society*, vol. 167, no. 3, p. 037 552, 2020 (cit. on p. 25).

[73] Z. Simeu-Abazi and C. Sassine, "Maintenance integration in manufacturing systems by using stochastic petri nets," *International journal of production research*, vol. 37, no. 17, pp. 3927–3940, 1999 (cit. on p. 25).

[74] R. Accorsi, R. Manzini, P. Pascarella, M. Patella, and S. Sassi, "Data mining and machine learning for condition-based maintenance," *Procedia Manufacturing*, vol. 11, pp. 1153–1161, 2017 (cit. on p. 25).

[75] B. Dowdeswell, R. Sinha, and S. G. MacDonell, "Finding faults: A scoping study of fault diagnostics for industrial cyber–physical systems," *Journal of Systems and Software*, vol. 168, p. 110 638, 2020 (cit. on p. 25).

[76] M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, 2018 (cit. on pp. 25, 107, 109, 113).

[77] Y.-g. Chen, "Applications of bayesian network in fault diagnosis of braking system," in *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, IEEE, vol. 1, 2011, pp. 234–237 (cit. on p. 25).

[78] P. Aivaliotis, K. Georgoulias, and K. Alexopoulos, "Using digital twin for maintenance applications in manufacturing: State of the art and gap analysis," in *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, 2019, pp. 1–5 (cit. on pp. 25, 109).

[79] M. Vathoopan, M. Johny, A. Zoitl, and A. Knoll, "Modular fault ascription and corrective maintenance using a digital twin," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1041–1046, 2018 (cit. on pp. 25, 109).

[80] P. Z. Schulte, "A state machine architecture for aerospace vehicle fault protection," Ph.D. dissertation, Georgia Institute of Technology, 2018 (cit. on p. 26).

[81] D. Novikov, R. V. Yampolskiy, and L. Reznik, "Anomaly detection based intrusion detection," in *Third International Conference on Information Technology: New Generations (ITNG'06)*, IEEE, 2006, pp. 420–425 (cit. on p. 26).

[82] M. Reif, M. Goldstein, A. Stahl, and T. M. Breuel, "Anomaly detection by combining decision trees and parametric densities," in *2008 19th International Conference on Pattern Recognition*, IEEE, 2008, pp. 1–4 (cit. on p. 26).

[83] M. Cakir, M. A. Guvenc, and S. Mistikoglu, "The experimental application of popular machine learning algorithms on predictive maintenance and the design of iiot based condition monitoring system," *Computers & Industrial Engineering*, vol. 151, p. 106 948, 2021 (cit. on pp. 26, 27).

[84] Q. Wang, W. Jiao, P. Wang, and Y. Zhang, "A tutorial on deep learning-based data analytics in manufacturing through a welding case study," *Journal of Manufacturing Processes*, vol. 63, pp. 2–13, 2021 (cit. on p. 26).

[85] J. Long, J. Mou, L. Zhang, S. Zhang, and C. Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots," *Journal of Manufacturing Systems*, 2020 (cit. on pp. 26, 29).

[86] S. Shakya and S. Sigdel, "An approach to develop a hybrid algorithm based on support vector machine and naive bayes for anomaly detection," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, IEEE, 2017, pp. 323–327 (cit. on p. 26).

[87] M.-C. Chiu, C.-D. Tsai, and T.-L. Li, "An integrative machine learning method to improve fault detection and productivity performance in a cyber-physical system," *Journal of Computing and Information Science in Engineering*, vol. 20, no. 2, 2020 (cit. on pp. 26, 53).

[88] M. Jung, O. Niculita, and Z. Skaf, "Comparison of different classification algorithms for fault detection and fault isolation in complex systems," *Procedia Manufacturing*, vol. 19, pp. 111–118, 2018 (cit. on pp. 26, 27).

[89] C. Giannetti and R. S. Ransing, "Risk based uncertainty quantification to improve robustness of manufacturing operations," *Computers & Industrial Engineering*, vol. 101, pp. 70–80, 2016 (cit. on p. 26).

[90] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001 (cit. on pp. 27, 41).

[91] J. S. Kim, J. Kim, and J. Y. Lee, "Die-casting defect prediction and diagnosis system using process condition data," *Procedia Manufacturing*, vol. 51, pp. 359–364, 2020 (cit. on p. 27).

[92] S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using iot data in real-time," *Expert Systems with Applications*, vol. 173, p. 114 598, 2021 (cit. on p. 27).

[93] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999 (cit. on p. 27).

[94] K. C. Gryllias and I. A. Antoniadis, "A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 326–344, 2012 (cit. on p. 27).

[95] S. Salcedo-Sanz, J. L. Rojo-Álvarez, M. Martínez-Ramón, and G. Camps-Valls, "Support vector machines in engineering: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 3, pp. 234–267, 2014 (cit. on p. 27).

[96] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992 (cit. on p. 27).

[97] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015 (cit. on pp. 28, 41, 53, 64).

[98] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, Springer, 2014, pp. 818–833 (cit. on p. 28).

[99] J. Zheng, Y. Dai, Y. Liang, Q. Liao, and H. Zhang, "An online real-time estimation tool of leakage parameters for hazardous liquid pipelines," *International Journal of Critical Infrastructure Protection*, vol. 31, p. 100 389, 2020 (cit. on p. 29).

[100] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017 (cit. on pp. 29, 51).

[101] J. Koskinen, M. Isohanni, H. Paajala, *et al.*, "How to use bibliometric methods in evaluation of scientific research? an example from finnish schizophrenia research," *Nordic journal of psychiatry*, vol. 62, no. 2, pp. 136–143, 2008 (cit. on p. 30).

[102] M. T. Amin, F. Khan, and M. J. Zuo, "A bibliometric analysis of process system failure and reliability literature," *Engineering Failure Analysis*, vol. 106, p. 104 152, 2019 (cit. on p. 30).

[103] E. Garfield, "Keywords plus-isi's breakthrough retrieval method. 1. expanding your searching power on current-contents on diskette," *Current contents*, vol. 32, pp. 5–9, 1990 (cit. on p. 30).

[104] J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan, "Comparing keywords plus of wos and author keywords: A case study of patient adherence research," *Journal of the Association for Information Science and Technology*, vol. 67, no. 4, pp. 967–972, 2016 (cit. on p. 30).

[105] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the American Society for information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011 (cit. on pp. 30, 32).

[106] G. Atzeni, G. Vignali, L. Tebaldi, and E. Bottani, "A bibliometric analysis on collaborative robots in logistics 4.0 environments," *Procedia Computer Science*, vol. 180, pp. 686–695, 2021 (cit. on p. 30).

[107] X. Wang, Z. Xu, and M. Škare, "A bibliometric analysis of economic research-ekonomska istraživanja (2007–2019)," *Economic Research-Ekonomska Istraživanja*, vol. 33, no. 1, pp. 865–886, 2020 (cit. on p. 30).

[108] M. J. Cobo, B Jürgens, V Herrero-Solana, M. A. Martínez, and E. Herrera-Viedma, "Industry 4.0: A perspective based on bibliometric analysis," *Procedia computer science*, vol. 139, pp. 364–371, 2018 (cit. on p. 30).

[109] Y. Riahi, T. Saikouk, A. Gunasekaran, and I. Badraoui, "Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions," *Expert Systems with Applications*, p. 114 702, 2021 (cit. on p. 30).

[110] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organizational Research Methods*, vol. 18, no. 3, pp. 429–472, 2015 (cit. on p. 32).

[111] M. F. Ab Razak, N. B. Anuar, R. Salleh, and A. Firdaus, "The rise of "malware": Bibliometric analysis of malware study," *Journal of Network and Computer Applications*, vol. 75, pp. 58–76, 2016 (cit. on p. 32).

[112] N. Ale Ebrahim, H. Salehi, M. A. Embi, *et al.*, "Effective strategies for increasing citation frequency," *International Education Studies*, vol. 6, no. 11, pp. 93–99, 2013 (cit. on p. 32).

[113] P. Mongeon and A. Paul-Hus, "The journal coverage of bibliometric databases: A comparison of scopus and web of science," *The journal coverage of Web of Science and Scopus: a comparative analysis. Available online: DOI*, vol. 10, 2014 (cit. on p. 32).

[114] J. Choi, S. Yi, and K. C. Lee, "Analysis of keyword networks in mis research and implications for predicting knowledge evolution," *Information & Management*, vol. 48, no. 8, pp. 371–381, 2011 (cit. on p. 32).

[115] G. Chen, L. Xiao, C.-p. Hu, and X.-q. Zhao, "Identifying the research focus of library and information science institutions in china with institution-specific keywords," *Scientometrics*, vol. 103, no. 2, pp. 707–724, 2015 (cit. on p. 32).

[116] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988 (cit. on p. 32).

[117] G. Chen and L. Xiao, "Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods," *Journal of Informetrics*, vol. 10, no. 1, pp. 212–223, 2016 (cit. on p. 32).

[118] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Transactions on industrial electronics*, vol. 55, no. 12, pp. 4109–4126, 2008 (cit. on pp. 41, 43, 53).

[119] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016 (cit. on pp. 41, 43, 53).

[120] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining (kdd'13)," *ACM, New York, NY, USA*, 2013 (cit. on pp. 41, 43).

[121] J. R. Windmiller and J. Wang, "Wearable electrochemical sensors and biosensors: A review," *Electroanalysis*, vol. 25, no. 1, pp. 29–46, 2013 (cit. on p. 43).

[122] K. Muller, "M., dornhege, g., krauledat, m., curio, g. and blankertz, b," *Machine learning for real-time single-trial EEG-analysis: From Brain-computer interfacing to mental state monitoring. Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 82–90, 2008 (cit. on p. 43).

[123] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *Journal of Sound and Vibration*, vol. 388, pp. 154–170, 2017 (cit. on p. 43).

[124] I. J. Bigio, S. G. Bown, G. M. Briggs, *et al.*, "Diagnosis of breast cancer using elastic-scattering spectroscopy: Preliminary clinical results," *Journal of biomedical optics*, vol. 5, no. 2, pp. 221–228, 2000 (cit. on p. 43).

[125] J. Verrelst, J. Muñoz, L. Alonso, *et al.*, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3," *Remote Sensing of Environment*, vol. 118, pp. 127–139, 2012 (cit. on p. 43).

[126] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018 (cit. on p. 43).

[127] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *Journal of Hydrology*, vol. 569, pp. 387–408, 2019 (cit. on p. 43).

[128] B. Berg, B. Cortazar, D. Tseng, *et al.*, "Cellphone-based hand-held microplate reader for point-of-care testing of enzyme-linked immunosorbent assays," *ACS nano*, vol. 9, no. 8, pp. 7857–7866, 2015 (cit. on p. 43).

[129] Z. Jin, Y. Sun, and A. C. Cheng, "Predicting cardiovascular disease from real-time electrocardiographic monitoring: An adaptive machine learning approach on a cell phone," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2009, pp. 6889–6892 (cit. on p. 43).

[130] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017 (cit. on p. 43).

[131] J. Gonzaga, L. A. C. Meleiro, C Kiang, and R. Maciel Filho, "Ann-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Computers & chemical engineering*, vol. 33, no. 1, pp. 43–49, 2009 (cit. on p. 43).

[132] J. He, S. Gong, Y. Yu, *et al.*, "Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major chinese cities," *Environmental pollution*, vol. 223, pp. 484–496, 2017 (cit. on p. 43).

[133] S. Michie, L. Yardley, R. West, K. Patrick, and F. Greaves, "Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop," *Journal of medical Internet research*, vol. 19, no. 6, e232, 2017 (cit. on p. 43).

[134] V. Botu and R. Ramprasad, "Adaptive machine learning framework to accelerate ab initio molecular dynamics," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1074–1083, 2015 (cit. on p. 43).

[135] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition (2015)," *arXiv preprint arXiv:1512.03385*, 2016 (cit. on pp. 41, 53).

[136] M. Greenacre and J. Blasius, *Multiple correspondence analysis and related methods*. CRC press, 2006 (cit. on p. 46).

[137] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field," *Journal of informetrics*, vol. 5, no. 1, pp. 146–166, 2011 (cit. on pp. 47, 49).

[138] D. Ivanov, C. S. Tang, A. Dolgui, D. Battini, and A. Das, "Researchers' perspectives on industry 4.0: Multi-disciplinary analysis and opportunities for operations management," *International Journal of Production Research*, vol. 59, no. 7, pp. 2055–2078, 2021 (cit. on p. 51).

[139] N. Tuptuk and S. Hailes, "Security of smart manufacturing systems," *Journal of manufacturing systems*, vol. 47, pp. 93–106, 2018 (cit. on p. 51).

[140] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8394–8402, 2019 (cit. on p. 51).

[141] T. Wang, "Hybrid decision making: When interpretable models collaborate with black-box models," *CoRR, abs/1802.04346*, 2018 (cit. on p. 52).

[142] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, "Fooling lime and shap: Adversarial attacks on post hoc explanation methods," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 180–186 (cit. on p. 52).

[143] S. M. Lundberg, G. Erion, H. Chen, *et al.*, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56–67, 2020 (cit. on p. 52).

[144] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016 (cit. on pp. 52, 71).

[145] B. Das, S. Pal, and S. Bag, "Weld quality prediction in friction stir welding using wavelet analysis," *The International Journal of Advanced Manufacturing Technology*, vol. 89, no. 1-4, pp. 711–725, 2017 (cit. on p. 53).

[146] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996 (cit. on p. 53).

[147] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997 (cit. on p. 53).

[148] K. Wang, Q. Shu, and Q. Tu, "Technostress under different organizational environments: An empirical investigation," *Computers in human behavior*, vol. 24, no. 6, pp. 3002–3013, 2008 (cit. on p. 53).

[149] A. Kaplan and M. Haenlein, *Digital transformation and disruption: On big data, blockchain, artificial intelligence, and other things*, 2019 (cit. on p. 54).

[150] P. Gao, T. Yu, Y. Zhang, J. Wang, and J. Zhai, "Vibration analysis and control technologies of hydraulic pipeline system in aircraft: A review," *Chinese Journal of Aeronautics*, 2020 (cit. on p. 58).

[151] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, "Artificial intelligence and real-time predictive maintenance in industry 4.0: A bibliometric analysis," *AI and Ethics*, pp. 1–25, 2022 (cit. on pp. 58, 59, 67, 97, 104, 107).

[152] H. Dong, Z. Wang, S. X. Ding, and H. Gao, "A survey on distributed filtering and fault detection for sensor networks," *Mathematical Problems in Engineering*, vol. 2014, 2014 (cit. on p. 59).

[153] M. Xu, J. M. David, S. H. Kim, *et al.*, "The fourth industrial revolution: Opportunities and challenges," *International journal of financial research*, vol. 9, no. 2, pp. 90–95, 2018 (cit. on p. 59).

[154] Q. Hao, Y. Xue, W. Shen, B. Jones, and J. Zhu, "A decision support system for integrating corrective maintenance, preventive maintenance, and condition-based maintenance," in *Construction Research Congress 2010: Innovation for Reshaping Construction Practice*, 2010, pp. 470–479 (cit. on p. 59).

[155] S. Sarkar, X. Jin, and A. Ray, "Data-driven fault detection in aircraft engines with noisy sensor measurements," *Journal of Engineering for Gas Turbines and Power*, vol. 133, no. 8, 2011 (cit. on p. 59).

[156] S. Blank, T. Pfister, and K. Berns, "Sensor failure detection capabilities in low-level fusion: A comparison between fuzzy voting and kalman filtering," in *2011 IEEE International Conference on Robotics and Automation*, IEEE, 2011, pp. 4974–4979 (cit. on p. 59).

[157] M. Bastuck, T Baur, and A. Schütze, "Fusing cyclic sensor data with different cycle length," in *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, IEEE, 2016, pp. 72–77 (cit. on p. 59).

[158] Y. Wilhelm, P. Reimann, W. Gauchel, and B. Mitschang, "Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models," *Procedia CIRP*, vol. 99, pp. 278–283, 2021 (cit. on p. 59).

[159] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognostics—a review of current paradigms and practices," *The International Journal of Advanced Manufacturing Technology*, vol. 28, no. 9, pp. 1012–1024, 2006 (cit. on p. 59).

[160] N. Enshaei and F. Naderkhani, "Application of deep learning for fault diagnostic in induction machine's bearings," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2019, pp. 1–7 (cit. on p. 59).

[161] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236*, 2016 (cit. on p. 59).

[162] Y. Bengio, Y. LeCun, *et al.*, "Scaling learning algorithms towards ai," *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41, 2007 (cit. on p. 59).

[163] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995 (cit. on p. 59).

[164] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital signal processing*, vol. 73, pp. 1–15, 2018 (cit. on p. 59).

[165] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020 (cit. on pp. 59, 68).

[166] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the ai: Informing design practices for explainable ai user experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15 (cit. on p. 59).

[167] V. Arya, R. K. Bellamy, P.-Y. Chen, *et al.*, "One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques," *arXiv preprint arXiv:1909.03012*, 2019 (cit. on p. 59).

[168] M. Langer, D. Oster, T. Speith, *et al.*, "What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research," *Artificial Intelligence*, vol. 296, p. 103 473, 2021 (cit. on p. 59).

[169] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using shap," *arXiv preprint arXiv:1903.02407*, 2019 (cit. on p. 59).

[170] K. M. Hancock and Q Zhang, "A hybrid approach to hydraulic vane pump condition monitoring and fault detection," *Transactions of the ASABE*, vol. 49, no. 4, pp. 1203–1211, 2006 (cit. on p. 61).

[171] P Muthuvel, B. George, and G. Ramadass, "Magnetic-capacitive wear debris sensor plug for condition monitoring of hydraulic systems," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9120–9127, 2018 (cit. on p. 61).

[172] Z. He, S. Wang, K. Wang, and K. Li, "Prognostic analysis based on hybrid prediction method for axial piston pump," in *IEEE 10th International Conference on Industrial Informatics*, IEEE, 2012, pp. 688–692 (cit. on p. 61).

[173] A. Moosavian, M. Khazaee, H. Ahmadi, M. Khazaee, and G. Najafi, "Fault diagnosis and classification of water pump using adaptive neuro-fuzzy inference system based on vibration signals," *Structural Health Monitoring*, vol. 14, no. 5, pp. 402–410, 2015 (cit. on p. 61).

[174] N. Helwig, E. Pignanelli, and A. Schütze, "Condition monitoring of a complex hydraulic system using multivariate statistics," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, IEEE, 2015, pp. 210–215 (cit. on pp. 61–63, 79, 93, 97).

[175] X. Zhao, K. Zhang, and Y. Chai, "A multivariate time series classification based multiple fault diagnosis method for hydraulic systems," in *2019 Chinese Control Conference (CCC)*, IEEE, 2019, pp. 6819–6824 (cit. on p. 61).

[176] T. Schneider, N. Helwig, and A. Schütze, "Automatic feature extraction and selection for classification of cyclical time series data," *tm-Technisches Messen*, vol. 84, no. 3, pp. 198–206, 2017 (cit. on p. 61).

[177] V. V. Shanbhag, T. J. Meyer, L. W. Caspers, and R. Schlanbusch, "Failure monitoring and predictive maintenance of hydraulic cylinder—state-of-the-art review," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 6, pp. 3087–3103, 2021 (cit. on p. 61).

[178] J. Zhu, D. He, and E. Bechhoefer, "Survey of lubrication oil condition monitoring, diagnostics, and prognostics techniques and systems," *Journal of chemical science and technology*, vol. 2, no. 3, pp. 100–115, 2013 (cit. on p. 61).

[179] S. Raadnui and S. Kleesuwan, "Low-cost condition monitoring sensor for used oil analysis," *Wear*, vol. 259, no. 7-12, pp. 1502–1506, 2005 (cit. on p. 61).

[180] S. Kumar, P. Mukherjee, and N. Mishra, "Online condition monitoring of engine oil," *Industrial lubrication and tribology*, 2005 (cit. on p. 61).

[181] T Bley, E Pignanelli, and A Schütze, "Multi-channel ir sensor system for determination of oil degradation," *Journal of Sensors and Sensor Systems*, vol. 3, no. 1, pp. 121–132, 2014 (cit. on p. 61).

[182] J. Bhat and B. Sonawane, "Condition monitoring of worm gearbox through oil analysis," in *Recent Trends in Engineering Design*, Springer, 2021, pp. 289–296 (cit. on p. 61).

[183] A. Y. Goharrizi and N. Sepehri, "A wavelet-based approach for external leakage detection and isolation from internal leakage in valve-controlled hydraulic actuators," *IEEE Transactions on industrial electronics*, vol. 58, no. 9, pp. 4374–4384, 2010 (cit. on p. 61).

[184] L. An and N. Sepehri, "Hydraulic actuator leakage fault detection using extended kalman filter," *International Journal of Fluid Power*, vol. 6, no. 1, pp. 41–51, 2005 (cit. on p. 61).

[185] A. C. Tan, P. S. Chua, and G. Lim, "Condition monitoring of a water hydraulic cylinder by vibration analysis," *Journal of testing and evaluation*, vol. 28, no. 6, pp. 507–512, 2000 (cit. on p. 61).

[186] H. Yunbo, G Lim, P Chua, and A Tan, "Monitoring the condition of loaded modern water hydraulic axial piston motor and cylinder," in *Proceedings of the Fifth International Conference on Fluid Power Transmission and Control*, Citeseer, 2001, pp. 447–451 (cit. on p. 61).

[187] V. V. Shanbhag, T. J. Meyer, L. W. Caspers, and R. Schlanbusch, "Condition monitoring of hydraulic cylinder seals using acoustic emissions," *The International Journal of Advanced Manufacturing Technology*, vol. 109, no. 5, pp. 1727–1739, 2020 (cit. on p. 61).

[188] M. Ramachandran and Z. Siddique, "A data-driven, statistical feature-based, neural network method for rotary seal prognostics," *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, vol. 2, no. 2, 2019 (cit. on p. 61).

[189] M. Ramachandran and Z. Siddique, "Statistical time domain feature based approach to assess the performance degradation of rotary seals," in *ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers, vol. 52187, 2018, V013T05A071 (cit. on p. 61).

[190] R. Corbally and A. Malekjafarian, "A data-driven approach for drive-by damage detection in bridges considering the influence of temperature change," *Engineering Structures*, vol. 253, p. 113 783, 2022 (cit. on p. 61).

[191] Y. Wen, M. F. Rahman, H. Xu, and T.-L. B. Tseng, "Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective," *Measurement*, vol. 187, p. 110 276, 2022 (cit. on p. 61).

[192] T. Li, S. Wang, E. Zio, J. Shi, and Z. Ma, "A numerical approach for predicting the remaining useful life of an aviation hydraulic pump based on monitoring abrasive debris generation," *Mechanical Systems and Signal Processing*, vol. 136, p. 106 519, 2020 (cit. on p. 61).

[193] E. Quatrini, F. Costantino, C. Pocci, and M. Tronci, "Predictive model for the degradation state of a hydraulic system with dimensionality reduction," *Procedia Manufacturing*, vol. 42, pp. 516–523, 2020 (cit. on pp. 62, 63).

[194] E. Georgievskaia, "Predictive analytics as a way to smart maintenance of hydraulic turbines," *Procedia Structural Integrity*, vol. 28, pp. 836–842, 2020 (cit. on p. 62).

[195] H. Zhang, H. Shi, W. Li, *et al.*, "A novel impedance micro-sensor for metal debris monitoring of hydraulic oil," *Micromachines*, vol. 12, no. 2, p. 150, 2021 (cit. on p. 62).

[196] C. König and A. M. Helmi, "Sensitivity analysis of sensors in a hydraulic condition monitoring system using cnn models," *Sensors*, vol. 20, no. 11, p. 3307, 2020 (cit. on pp. 62, 63, 88).

[197] K. Kim and J. Jeong, "Deep learning-based data augmentation for hydraulic condition monitoring system," *Procedia Computer Science*, vol. 175, pp. 20–27, 2020 (cit. on p. 62).

[198] S. S. Chawathe, "Condition monitoring of hydraulic systems by classifying sensor data streams," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2019, pp. 0898–0904 (cit. on pp. 63, 80).

[199] D. Z. Fawwaz and S.-H. Chung, "Real-time and robust hydraulic system fault detection via edge computing," *Applied Sciences*, vol. 10, no. 17, p. 5933, 2020 (cit. on p. 63).

[200] Y. Lei, W. Jiang, A. Jiang, Y. Zhu, H. Niu, and S. Zhang, "Fault diagnosis method for hydraulic directional valves integrating pca and xgboost," *Processes*, vol. 7, no. 9, p. 589, 2019 (cit. on p. 63).

[201] Y. Bengio, "Deep learning of representations: Looking forward," in *International Conference on Statistical Language and Speech Processing*, Springer, 2013, pp. 1–37 (cit. on pp. 64, 108).

[202] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017 (cit. on pp. 64, 108).

[203] A. Z. Woldaregay, E. Årsand, T. Botsis, D. Albers, L. Mamykina, and G. Hartvigsen, "Data-driven blood glucose pattern classification and anomalies detection: Machine-learning applications in type 1 diabetes," *Journal of medical Internet research*, vol. 21, no. 5, e11030, 2019 (cit. on p. 64).

[204] T. Szandała, "Review and comparison of commonly used activation functions for deep neural networks," in *Bio-inspired neurocomputing*, Springer, 2021, pp. 203–224 (cit. on p. 64).

[205] A. M. Alhassan and W. M. N. W. Zainon, "Brain tumor classification in magnetic resonance image using hard swish-based relu activation function-convolutional neural network," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9075–9087, 2021 (cit. on p. 64).

[206] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, IEEE, 2018, pp. 1–2 (cit. on p. 65).

[207] V. Dignum, *Ethics in artificial intelligence: Introduction to the special issue*, 2018 (cit. on p. 66).

[208] E. Turiel, *The culture of morality: Social development, context, and conflict*. Cambridge University Press, 2002 (cit. on p. 66).

[209] I. Rahwan, "Society-in-the-loop: Programming the algorithmic social contract," *Ethics and Information Technology*, vol. 20, no. 1, pp. 5–14, 2018 (cit. on p. 66).

[210] P. Vamplew, R. Dazeley, C. Foale, S. Firmin, and J. Mummery, "Human-aligned artificial intelligence is a multiobjective problem," *Ethics and Information Technology*, vol. 20, no. 1, pp. 27–40, 2018 (cit. on p. 66).

[211] R. Belk, "Ethical issues in service robotics and artificial intelligence," *The Service Industries Journal*, pp. 1–17, 2020 (cit. on p. 67).

[212] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," *Ecological modelling*, vol. 160, no. 3, pp. 249–264, 2003 (cit. on p. 67).

[213] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Recognition*, vol. 65, pp. 211–222, 2017 (cit. on p. 67).

[214] A. Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *Big data*, vol. 5, no. 2, pp. 153–163, 2017 (cit. on p. 67).

[215] R. Goebel, A. Chander, K. Holzinger, *et al.*, "Explainable ai: The new 42?" In *International cross-domain conference for machine learning and knowledge extraction*, Springer, 2018, pp. 295–303 (cit. on p. 67).

[216] M. Harbers, K. van den Bosch, and J.-J. Meyer, "Design and evaluation of explainable bdi agents," in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE, vol. 2, 2010, pp. 125–132 (cit. on p. 67).

[217] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, 2019 (cit. on p. 67).

[218] D. Erhan, A. Courville, and Y. Bengio, "Understanding representations learned in deep architectures," *Department dInformatique et Recherche Operationnelle, University of Montreal, QC, Canada, Tech. Rep*, vol. 1355, no. 1, 2010 (cit. on p. 69).

[219] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, e0130140, 2015 (cit. on p. 70).

[220] S.-K. Yeom, P. Seegerer, S. Lapuschkin, *et al.*, "Pruning by explaining: A novel criterion for deep neural network pruning," *Pattern Recognition*, vol. 115, p. 107 899, 2021 (cit. on p. 70).

[221] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013 (cit. on p. 70).

[222] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International Conference on Machine Learning*, PMLR, 2017, pp. 3145–3153 (cit. on p. 70).

[223] A. Shrikumar, P. Greenside, A. Shcherbina, and A. Kundaje, "Not just a black box: Learning important features through propagating activation differences," *arXiv preprint arXiv:1605.01713*, 2016 (cit. on p. 71).

[224] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (xai): A survey," *arXiv preprint arXiv:2006.11371*, 2020 (cit. on pp. 71, 73).

[225] T. Peltola, "Local interpretable model-agnostic explanations of bayesian predictive models via kullback-leibler projections," *arXiv preprint arXiv:1810.02678*, 2018 (cit. on p. 72).

[226] S. Bramhall, H. Horn, M. Tieu, and N. Lohia, "Qlime-a quadratic local interpretable model-agnostic explanation approach," *SMU Data Science Review*, vol. 3, no. 1, p. 4, 2020 (cit. on p. 72).

[227] S. Shi, X. Zhang, and W. Fan, "A modified perturbed sampling method for local interpretable model-agnostic explanation," *arXiv preprint arXiv:2002.07434*, 2020 (cit. on p. 72).

[228] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017 (cit. on pp. 72–74).

[229] L. Bouneder, Y. Léo, and A. Lachapelle, "X-shap: Towards multiplicative explainability of machine learning," *arXiv preprint arXiv:2006.04574*, 2020 (cit. on p. 73).

[230] Z. T. Fernando, J. Singh, and A. Anand, "A study on the interpretability of neural retrieval models using deepshap," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, pp. 1005–1008 (cit. on p. 73).

[231] J. Yang, "Fast treeshap: Accelerating shap value computation for trees," *arXiv preprint arXiv:2109.09847*, 2021 (cit. on p. 73).

[232] M. V. García and J. L. Aznarte, "Shapley additive explanations for no2 forecasting," *Ecological Informatics*, vol. 56, p. 101 039, 2020 (cit. on p. 73).

[233] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020 (cit. on p. 73).

[234] C. Angeli, "An online expert system for fault diagnosis in hydraulic systems," *Expert systems*, vol. 16, no. 2, pp. 115–120, 1999 (cit. on p. 79).

[235] D. Zhang, W. Li, Y. Lin, and J. Bao, "An overview of hydraulic systems in wave energy application in china," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 7, pp. 4522–4526, 2012 (cit. on p. 79).

[236] A. Steinboeck, W. Kemmetmüller, C. Lassl, and A. Kugi, "Model-based condition monitoring of an electro-hydraulic valve," *Journal of Dynamic Systems, Measurement, and Control*, vol. 135, no. 6, 2013 (cit. on pp. 79, 80).

[237] N. Helwig, E Pignanelli, and A Schütze, "D8. 1-detecting and compensating sensor faults in a hydraulic condition monitoring system," *Proceedings SENSOR 2015*, pp. 641–646, 2015 (cit. on p. 79).

[238] C. C. Aggarwal, "Outlier analysis," in *Data mining*, Springer, 2015, pp. 237–263 (cit. on p. 103).

[239] W. Thomas, "Friction stir butt welding," *Int. Patent No. PCT/GB92/02203*, 1991 (cit. on p. 103).

[240] M. A. Sutton, B Yang, A. P. Reynolds, and R Taylor, "Microstructural studies of friction stir welds in 2024-t3 aluminum," *Materials science and engineering: A*, vol. 323, no. 1-2, pp. 160–166, 2002 (cit. on pp. 103, 111).

[241] K. Kumar and S. V. Kailas, "The role of friction stir welding tool on material flow and weld formation," *Materials Science and Engineering: A*, vol. 485, no. 1-2, pp. 367–374, 2008 (cit. on pp. 103, 111).

[242] T Jene, G Dobmann, G Wagner, and D Eifler, "Monitoring of the friction stir welding process to describe parameter effects on joint quality," *Welding in the World*, vol. 52, no. 9-10, pp. 47–53, 2008 (cit. on pp. 103, 111).

[243] D. Mishra, R. B. Roy, S. Dutta, S. K. Pal, and D. Chakravarty, "A review on sensor based monitoring and control of friction stir welding process and a roadmap to industry 4.0," *Journal of Manufacturing Processes*, vol. 36, pp. 373–397, 2018 (cit. on pp. 103, 106).

[244] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, "Health condition monitoring of a complex hydraulic system using deep neural network and deepshap explainable xai," *Advances in Engineering Software*, vol. 175, p. 103 339, 2023 (cit. on p. 103).

[245] X. He, F. Gu, and A. Ball, "A review of numerical analysis of friction stir welding," *Progress in Materials Science*, vol. 65, pp. 1–66, 2014 (cit. on p. 103).

[246] A. Mishra, "Artificial intelligence algorithms for the analysis of mechanical property of friction stir welded joints by using python programming," *Welding Technology Review*, vol. 92, no. 6, pp. 7–16, 2020 (cit. on pp. 103, 104).

[247] D. Kleiner and C. R. Bird, "Signal processing for quality assurance in friction stir welds," *Insight-Non-Destructive Testing and Condition Monitoring*, vol. 46, no. 2, pp. 85–87, 2004 (cit. on p. 103).

[248] R. S. Mishra and Z. Ma, "Friction stir welding and processing," *Materials science and engineering: R: reports*, vol. 50, no. 1-2, pp. 1–78, 2005 (cit. on pp. 103, 110).

[249] X. Zhu and Y. Chao, "Numerical simulation of transient temperature and residual stresses in friction stir welding of 304l stainless steel," *Journal of materials processing technology*, vol. 146, no. 2, pp. 263–272, 2004 (cit. on pp. 103, 104).

[250] G Yoshikawa, F Miyasaka, Y Hirata, Y Katayama, and T Fuse, "Development of numerical simulation model for fsw employing particle method," *Science and Technology of Welding and Joining*, vol. 17, no. 4, pp. 255–263, 2012 (cit. on pp. 103, 106, 107).

[251] P. Colegrove and H. Shercliff, "Cfd modelling of friction stir welding of thick plate 7449 aluminium alloy," *Science and Technology of welding and joining*, vol. 11, no. 4, pp. 429–441, 2006 (cit. on p. 103).

[252] H Schmidt, J. Hattel, and J. Wert, "An analytical model for the heat generation in friction stir welding," *Modelling and simulation in materials science and engineering*, vol. 12, no. 1, p. 143, 2003 (cit. on p. 103).

[253] T. Seidel and A. P. Reynolds, "Two-dimensional friction stir welding process model based on fluid mechanics," *Science and technology of welding and joining*, vol. 8, no. 3, pp. 175–183, 2003 (cit. on p. 103).

[254] R Nandan, G. Roy, T. Lienert, and T DebRoy, "Numerical modelling of 3d plastic flow and heat transfer during friction stir welding of stainless steel," *Science and Technology of Welding and Joining*, vol. 11, no. 5, pp. 526–537, 2006 (cit. on p. 103).

[255] A. Tongne, C. Desrayaud, M. Jahazi, and E. Feulvarch, "On material flow in friction stir welded al alloys," *Journal of Materials Processing Technology*, vol. 239, pp. 284–296, 2017 (cit. on pp. 103, 106, 111).

[256] Z Zhang and H. Zhang, "Numerical studies on effect of axial pressure in friction stir welding," *Science and Technology of Welding and Joining*, vol. 12, no. 3, pp. 226–248, 2007 (cit. on pp. 104, 111).

[257] S. J. Russell, *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010 (cit. on p. 104).

[258] U. Chadha, S. K. Selvaraj, N. Gunreddy, *et al.*, "A survey of machine learning in friction stir welding, including unresolved issues and future research directions," *Material Design & Processing Communications*, vol. 2022, 2022 (cit. on p. 104).

[259] R. Sandeep and A. Natarajan, "Prediction of peak temperature value in friction lap welding of aluminium alloy 7475 and pps polymer hybrid joint using machine learning approaches," *Materials Letters*, vol. 308, p. 131 253, 2022 (cit. on p. 104).

[260] A. Mishra, "Machine learning approach for defects identification in dissimilar friction stir welded aluminium alloys aa 7075-aa 1100 joints," *Journal of Aircraft and Spacecraft Technology*, vol. 4, no. 1, pp. 88–95, 2020 (cit. on p. 104).

[261] X. Jia, J. Willard, A. Karpatne, *et al.*, "Physics guided rnns for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proceedings of the 2019 SIAM International Conference on Data Mining*, SIAM, 2019, pp. 558–566 (cit. on pp. 104, 121).

[262] B. Meyghani, M. B. Awang, S. S. Emamian, M. K. B. Mohd Nor, and S. R. Pedapati, "A comparison of different finite element methods in the thermal analysis of friction stir welding (fsw)," *Metals*, vol. 7, no. 10, p. 450, 2017 (cit. on pp. 106, 107).

[263] A. K. Kadian and P. Biswas, "The study of material flow behaviour in dissimilar material fsw of aa6061 and cu-b370 alloys plates," *Journal of Manufacturing Processes*, vol. 34, pp. 96–105, 2018 (cit. on p. 106).

[264] A. Tartakovsky, G. Grant, X. Sun, and M. Khaleel, "Modeling of friction stir welding (fsw) process with smooth particle hydrodynamics (sph)," SAE Technical Paper, Tech. Rep., 2006 (cit. on p. 106).

[265] K. Fraser, L. St-Georges, and L. I. Kiss, "A mesh-free solid-mechanics approach for simulating the friction stir-welding process," *Joining Technologies*, pp. 27–52, 2016 (cit. on p. 106).

[266] S. Nasiri, M. R. Khosravani, and K. Weinberg, "Fracture mechanics and mechanical fault detection by artificial intelligence methods: A review," *Engineering Failure Analysis*, vol. 81, pp. 270–293, 2017 (cit. on p. 106).

[267] N. Ghetiya and K. Patel, "Prediction of tensile strength in friction stir welded aluminium alloy using artificial neural network," *Procedia Technology*, vol. 14, pp. 274–281, 2014 (cit. on pp. 106, 108).

[268] S. Verma, M. Gupta, and J. P. Misra, "Performance evaluation of friction stir welding using machine learning approaches," *MethodsX*, vol. 5, pp. 1048–1058, 2018 (cit. on pp. 107, 108).

[269] M. Kumar and N. Yadav, "Multilayer perceptrons and radial basis function neural network methods for the solution of differential equations: A survey," *Computers & Mathematics with Applications*, vol. 62, no. 10, pp. 3796–3811, 2011 (cit. on p. 107).

[270] A. Shrivastava, F. E. Pfefferkorn, N. A. Duffie, *et al.*, "Physics-based process model approach for detecting discontinuity during friction stir welding," *The International Journal of Advanced Manufacturing Technology*, vol. 79, no. 1, pp. 605–614, 2015 (cit. on p. 107).

[271] V. Malik, N. Sanjeev, H. S. Hebbar, and S. V. Kailas, "Investigations on the effect of various tool pin profiles in friction stir welding using finite element simulations," *Procedia Engineering*, vol. 97, pp. 1060–1068, 2014 (cit. on p. 107).

[272] H Schmidt and J. Hattel, "Modelling heat flow around tool probe in friction stir welding," *Science and Technology of Welding and joining*, vol. 10, no. 2, pp. 176–186, 2005 (cit. on p. 107).

[273] M. Atwya and G. Panoutsos, "Transient thermography for flaw detection in friction stir welding: A machine learning approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4423–4435, 2019 (cit. on p. 108).

[274] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999 (cit. on p. 108).

[275] J. R. Quinlan *et al.*, "Bagging, boosting, and c4. 5," in *AAAI/IAAI, Vol. 1*, 1996, pp. 725–730 (cit. on p. 108).

[276] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996 (cit. on p. 108).

[277] G. Capuano and J. J. Rimoli, "Smart finite elements: A novel machine learning application," *Computer Methods in Applied Mechanics and Engineering*, vol. 345, pp. 363–381, 2019 (cit. on p. 108).

[278] D. M. Hawkins, "The problem of overfitting," *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1–12, 2004 (cit. on p. 108).

[279] J. H. Friedman, "On bias, variance, 0/1—loss, and the curse-of-dimensionality," *Data mining and knowledge discovery*, vol. 1, no. 1, pp. 55–77, 1997 (cit. on p. 108).

[280] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM computing surveys (CSUR)*, vol. 27, no. 3, pp. 326–327, 1995 (cit. on p. 108).

[281] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016 (cit. on p. 108).

[282] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013 (cit. on p. 108).

[283] G. E. Hinton, T. J. Sejnowski, *et al.*, "Learning and relearning in boltzmann machines," *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, no. 282-317, p. 2, 1986 (cit. on p. 108).

[284] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research [research frontier]," *IEEE computational intelligence magazine*, vol. 5, no. 4, pp. 13–18, 2010 (cit. on p. 108).

[285] H. Okuyucu, A. Kurt, and E. Arcaklioglu, "Artificial neural network application to the friction stir welding of aluminum plates," *Materials & design*, vol. 28, no. 1, pp. 78–84, 2007 (cit. on p. 108).

[286] M. Stoffel, F. Bamer, and B. Markert, "Artificial neural networks and intelligent finite elements in non-linear structural mechanics," *Thin-Walled Structures*, vol. 131, pp. 102–106, 2018 (cit. on p. 108).

[287] S. Oh and H. Ki, "Deep learning model for predicting hardness distribution in laser heat treatment of aisi h13 tool steel," *Applied Thermal Engineering*, vol. 153, pp. 583–595, 2019 (cit. on p. 108).

[288] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996 (cit. on p. 108).

[289] B. Eren, M. A. Guvenc, and S. Mistikoglu, "Artificial intelligence applications for friction stir welding: A review," *Metals and Materials International*, vol. 27, no. 2, pp. 193–219, 2021 (cit. on p. 108).

[290] C Patel, S. Das, and R. G. Narayanan, "Cafe modeling, neural network modeling, and experimental investigation of friction stir welding," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 227, no. 6, pp. 1164–1176, 2013 (cit. on p. 108).

[291] C. J. Arthurs and A. P. King, "Active training of physics-informed neural networks to aggregate and interpolate parametric solutions to the navier-stokes equations," *Journal of Computational Physics*, vol. 438, p. 110 364, 2021 (cit. on p. 109).

[292] X. Jin, S. Cai, H. Li, and G. E. Karniadakis, "Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations," *Journal of Computational Physics*, vol. 426, p. 109 951, 2021 (cit. on p. 109).

[293] K. Zubov, Z. McCarthy, Y. Ma, *et al.*, "Neuralpde: Automating physics-informed neural networks (pinns) with error approximations," *arXiv preprint arXiv:2107.09443*, 2021 (cit. on p. 109).

[294] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (pinns) for fluid mechanics: A review," *Acta Mechanica Sinica*, pp. 1–12, 2022 (cit. on p. 109).

[295] Z. Fang and J. Zhan, "Deep physical informed neural networks for metamaterial design," *IEEE Access*, vol. 8, pp. 24 506–24 513, 2019 (cit. on p. 109).

[296] S. Cai, Z. Wang, S. Wang, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks for heat transfer problems," *Journal of Heat Transfer*, vol. 143, no. 6, 2021 (cit. on p. 109).

[297] L. Yang, X. Meng, and G. E. Karniadakis, "B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data," *Journal of Computational Physics*, vol. 425, p. 109 913, 2021 (cit. on p. 109).

[298] A. A. Ramabathiran and P. Ramachandran, "Spinn: Sparse, physics-based, and partially interpretable neural networks for pdes," *Journal of Computational Physics*, vol. 445, p. 110 600, 2021 (cit. on p. 109).

[299] G. Pang, L. Lu, and G. E. Karniadakis, "Fpinns: Fractional physics-informed neural networks," *SIAM Journal on Scientific Computing*, vol. 41, no. 4, A2603–A2626, 2019 (cit. on p. 109).

[300] Y. Guo, X. Cao, B. Liu, and M. Gao, "Solving partial differential equations using deep learning and physical constraints," *Applied Sciences*, vol. 10, no. 17, p. 5917, 2020 (cit. on p. 109).

[301] D. Lucor, A. Agrawal, and A. Sergent, "Physics-aware deep neural networks for surrogate modeling of turbulent natural convection," *arXiv preprint arXiv:2103.03565*, 2021 (cit. on p. 109).

[302] X. Liu and M. Almekkawy, "Ultrasound computed tomography using physical-informed neural network," in *2021 IEEE International Ultrasonics Symposium (IUS)*, IEEE, 2021, pp. 1–4 (cit. on p. 109).

[303] X. Li and W. Zhang, "Physics-informed deep learning model in wind turbine response prediction," *Renewable Energy*, vol. 185, pp. 932–944, 2022 (cit. on p. 109).

[304] A Squillace, A De Fenzo, G Giorleo, and F Bellucci, "A comparison between fsw and tig welding techniques: Modifications of microstructure and pitting corrosion resistance in aa 2024-t3 butt joints," *Journal of Materials Processing Technology*, (cit. on p. 110).

[305] K. Fraser, s.-g. Lyne, and L. Kiss, "A mesh-free solid-mechanics approach for simulating the friction stir-welding process," in Sep. 2016. DOI: 10.5772/64159 (cit. on p. 110).

[306] G. Chen, Q. Ma, S. Zhang, J. Wu, G. Zhang, and Q. Shi, "Computational fluid dynamics simulation of friction stir welding: A comparative study on different frictional boundary conditions," *Journal of Materials Science & Technology*, vol. 34, no. 1, pp. 128–134, 2018 (cit. on pp. 111–113).

[307] E. Feulvarch, J.-C. Roux, and J.-M. Bergheau, "A simple and robust moving mesh technique for the finite element simulation of friction stir welding," *Journal of Computational and Applied Mathematics*, vol. 246, pp. 269–277, 2013 (cit. on p. 112).

[308] P. Bussetta, É. Feulvarch, A. Tongne, R. Boman, J.-M. Bergheau, and J.-P. Ponthot, "Two 3d thermomechanical numerical models of friction stir welding processes with a trigonal pin," *Numerical Heat Transfer, Part A: Applications*, vol. 70, no. 9, pp. 995–1008, 2016 (cit. on pp. 112, 113).

[309] P. A. Colegrove and H. R. Shercliff, "3-dimensional cfd modelling of flow round a threaded friction stir welding tool profile," *Journal of materials processing technology*, vol. 169, no. 2, pp. 320–327, 2005 (cit. on p. 113).

[310] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *towards data science*, vol. 6, no. 12, pp. 310–316, 2017 (cit. on p. 118).

[311] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural networks*, vol. 6, no. 6, pp. 861–867, 1993 (cit. on p. 118).

[312] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML*, 2010, pp. 21–7 (cit. on p. 119).

[313] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018 (cit. on p. 119).

[314] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014 (cit. on p. 119).

[315] P. Gallinari and T. Cibas, "Practical complexity control in multilayer perceptrons," *Signal Processing*, vol. 74, no. 1, pp. 29–46, 1999 (cit. on p. 120).

[316] B. Neyshabur, R. Tomioka, and N. Srebro, "In search of the real inductive bias: On the role of implicit regularization in deep learning," *arXiv preprint arXiv:1412.6614*, 2014 (cit. on p. 121).

[317] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017 (cit. on p. 121).

[318] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural computation*, vol. 4, no. 1, pp. 1–58, 1992 (cit. on p. 121).

[319] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014 (cit. on pp. 121, 122).

[320] P. L. Bartlett, "For valid generalization the size of the weights is more important than the size of the network," in *Advances in neural information processing systems*, 1997, pp. 134–140 (cit. on p. 122).

[321] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006 (cit. on p. 126).