Fault Diagnosis in a Permanent Magnet Synchronous Motor using Deep Learning

Von der Fakultät für Elektrotechnik
der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg zur Erlangung
des akademischen Grades eines Doktor-Ingenieurs
genehmigte

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurswissenschaften

Vorgelegt von

Pedram Quseiri Darbandeh

aus Sanandaj

Hamburg 2023

Gutachter:

Univ.-Prof. Dr.-Ing. Christian Kreischer
Helmut-Schmidt-Universität,Universität der Bundeswehr Hamburg

Univ.-Prof. Dr. rer. nat. Marcus Stiemer
Helmut-Schmidt-Universität,Universität der Bundeswehr Hamburg

Vorsitzender:

Univ.-Prof. Dr.-Ing. Joachim Horn
Helmut-Schmidt-Universität,Universität der Bundeswehr Hamburg

Tag der mündlichen Prüfung: 03.02.2023

ACKNOWLEDGEMENTS

I am deeply grateful to my parents for their unwavering support, encouragement, and love throughout my academic journey. Their guidance and inspiration have been invaluable, and I cannot thank them enough.

I extend my heartfelt appreciation to Professor Dr. Kreischer and the EMA research group for their invaluable guidance, mentorship, and support. Collaborating with this esteemed group of researchers has been a privilege, providing me with opportunities to learn and grow from their expertise. I am honored to have been a part of their team, and grateful to Professor Dr. Kreischer for his supervision and guidance in developing this thesis.

I am also sincerely grateful to Professor Dr. Stiemer and Professor Dr. Horn for taking the time to read and evaluate my thesis. I especially want to thank Professor Dr. Stiemer for his insightful feedback that has been instrumental in shaping my research. I owe a debt of gratitude to my former professors, Professor Dr. Ardebili and Professor Dr. Aliyari Shooredeli, for their mentorship and encouragement during my studies. They provided me with a strong foundation for my future academic pursuits.

I am thankful for all the participants who generously shared their time and insights with me throughout my research. Without their contributions, this thesis would not have been possible.

Finally, I am humbled and grateful for the support, guidance, and encouragement of so many individuals and institutions. This thesis is dedicated to my parents, who have been my guiding light throughout my academic journey. Thank you all from the bottom of my heart for your invaluable contributions to my success.

ABSTRACT

Condition Monitoring (CM) of electrical equipment, especially electrical machines, has an important role in enhancing reliability and preventing disturbance in production centres such as power plants and factories. One of the important steps of CM's process is diagnosing faults at the proper time to prevent the defect from spreading to other components of the system. Nowadays, the Permanent Magnet Synchronous Machine (PMSM) which uses permanent magnets instead of winding in its structure is widely used among other types of electrical machines. This thesis proposes a new algorithm based on Variational Autoencoder (VAE) as an unsupervised deep learning method to detect different faults with different severities. In this research work, current, vibration and sound signals are collected from a 1 kW prototyped PMSM. This motor is simulated in Finite Element Analysis (FEA) software and compared with the prototyped one. Demagnetization, bearing, and eccentricity faults are selected to be implemented on a prototyped PMSM. The performance of VAE is evaluated for different operating conditions consisting of different loads and speeds. Also, the VAE is investigated for classifying data with different complexity. The feature extraction methods which are known as the pre-processing methods can improve the performance of VAE by reducing dimensionality and extracting important data features. Besides, these methods improve the network's training time by reducing input data size. In this thesis, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used as feature extraction methods. The performance of VAE-PCA and VAE-LDA is compared, and it is shown that VAE-PCA has a slightly better performance than VAE-LDA in most operating conditions, especially in the high-class number in which the complexity is higher compared to other cases. Also, this method is developed for analysing different combinations of sensors. Results show that the combination of current and sound signals can be a successful way to diagnose faults with and without pre-processing methods. In addition, the performance of VAE is compared with other machine learning methods such as Naïve Bayes (NB) and K-Nearest Neighbourhood (KNN) with and without pre-processing methods from accuracy and training time aspects. It is demonstrated that the performance of VAE for classifying the low and high number of classes of data is proper from accuracy and training time aspects. However, the performance of machine learning methods in the high number of classes is not proper for training time.

TABLE OF CONTENTS

SYMBOLS AND INDICES

Symbol	Description	Unit	Symbol	Description	Unit
a_{i1}	The ith component of a_1	-	$h(x)$	The Variance of distribution	-
$a_{i,k}$	The component of a matrices in the row ith and columns kth	-	H	Hypothesis space	-
a_k	The kth component of transformation	-	H	Magnetic field strength	A/M
A_{dem}	Amplitude of voltage related to demagnetization	V	$H\mathrm{c}$	Coercivity of magnet	H/m
A_{g}	Area of air gap	m^2	H_{c}	Magnetic field strength of core	H/m
A_{m}	Area of magnet	m^2	H_{g}	Magnetic field strength of air gap	H/m
B	Magnetic field density	T	H_{m}	Magnetic field strength of magnet	H/m
B_{g}	Magnetic field density in air gap	T	i	Current	A
B_{m}	Magnetic field density in magnet	T	i_{a}	Current phase a	A
B_{r}	Residual magnetic density	T	i_{b}	Current phase b	A
d	The distance between the center and the point k	-	i_{c}	Current phase c	A
D	Duty cycle	-	i_{d}	Current in d-axis	A
D	The training space	-	I_{dc}	DC current	A
e_{slot}	Induced voltage in the slot	V	i_{q}	Current in q-axis	A
E	Induce voltage	V	k	Effective coefficient for the flux	-
f	Activation function	-	K	Demagnetization severity	-
f_{dem}	Frequencies related to Demagnetization	Hz	K_{dem}	Demagnetization severity	-
$f(z)$	Deterministic function	-	$\widehat{K_t}$	The average torque [constant]	-
$g(x)$	The mean of distribution	-	L	The Lagrange function	-
G	Transform function from p to d features	-	l_{c}	Length path for flux in the core	m
h_{MAP}	Maximum a posteriori hypothesis	-	l_{cs1}	The flux path length of section (1) in the stator	m

Symbol	Description	Unit	Symbol	Description	Unit
l_{cs2}	The flux path length of section (2) in the stator	m	$\mathrm{P}(D\|h)$	The probability of D given h	-
l_{cs3}	The flux path length of section (3) in the stator	m	$\mathrm{P}(h)$	Prior probability of hypothesis h	-
l_{cr1}	The flux path length of section (1) in the rotor	m	$\mathrm{P}(D)$	Prior probability of training data D	-
l_{cr2}	The flux path length of section (2) in the rotor	m	$\mathrm{P}(z)$	Standard gaussian distribution	-
l_{cr3}	The flux path length of section (3) in the rotor	m	p	Number of pole paires	-
L_{d}	Inductance in d-axis	H	p	Number of features	-
l_{m}	Length path for flux in the magnet	m	$\mathrm{q}_x(z)$	Estimated distribution	-
l_{m1}	The flux path length of section (1) in the magnet	m	R_{g}	Estimated radius of rotation points around the center of mass	m
l_{m2}	The flux path length of section (2) in the magnet	m	R_{a}	Resistance in Phase a	Ω
l_{g}	Length path for flux in the air gap	m	R_{b}	Resistance in phase b	Ω
l_{g1}	The flux path length of section (1) in the air gap	m	R_{c}	Resistance in phase c	Ω
l_{g2}	The flux path length of section (2) in the air gap	m	R_{s}	Resistance of stator	Ω
L_{q}	Inductance in q-axis	H	$\mathrm{S_w}$	Within classes matrices	-
M	The number of observations	-	$\mathrm{S_B}$	Between classes matrices	-
M	Size of hidden layer	-	S	The scatter matrices	-
M_{ij}	The member of confusion matrix	-	$\mathrm{T_e}$	Electromagnetics torque	N.m
$M_{i,j}$	Mutual inductance of winding i and j	H	u_{d}	Voltage in d-axis	V
n	Number of row	-	u_{q}	Voltage in q-axis	V
N	Number of turns	-	U	Eigenvalues	-
N	Normal distribution	-	V	Eigenvector	-
$\mathrm{P}(h\|D)$	The probability of h given D	-	$V_{0,\mathrm{c}}$	The zero sequence of voltage	V

Symbol	Description	Unit	Symbol	Description	Unit
$V_{0,\mathrm{m}}$	The zero sequence of voltage related to the PM	V	$\overline{Z_k}$	The mean of kth principal component	-
V_{slot}	Induced voltage in slot	V	$z(j)$	The value for jth neuron	-
V_{MAP}	Maximum a posteriori for new instances	-	μ_0	The magnet permeability of free space	H/M
V_{a}	Voltage phase a	V	μ_{r}	The relative permeability	-
V_{b}	Voltage phase b	V	λ_{a}	Flux linkage in phase a	Wb
V_{c}	Voltage phase c	V	λ_{b}	Flux linkage in phase b	Wb
W^T	Transformation vector to project data on the new line	-	λ_{c}	Flux linkage in phase c	Wb
w_i	The ith coefficient value	-	λ_{pm}	Flux linkage of permanent magnet	Wb
$w_{j,i}$	The weight coefficient for input ($x(i)$) and neuron ($z(j)$)	-	λ_k	The kth eigenvalues	-
w_{i+1}	The new iteration's coefficient value	-	λ_{slot}	Flux linkage in slot	H
$w_{j,0}$	The bias coefficient of $z(j)$	-	$\lambda_{\mathrm{pm,v}}$	The vth harmonics of Flux linkage for the PM	H
$x_{i,j}$	The data set for row I and column j	-	ζ	Gaussian Distribution	-
x_i	The ith of input data	-	φ_{m}	Flux in the magnet	Wb
x	The input data	-	φ_{g}	Flux in the air gap	Wb
$\hat{x}$	The reconstructed data	-	μ_i	The mean of x	-
x_n	The nth component of input data	-	$\widetilde{\mu_i}$	The mean of y	-
$\bar{X}$	Mean of X	-	μ_x	Mean of estimated distribution	-
x_{ij}	The ith input data corresponding to jth Feature	-	ω	Angular velocity	Rad/s
y	The real target value	-	σ_x	Variance of estimated distribution	-
$\acute{y}$	The estimated of target value	-	$\sum$	Covariance Matrices	-
y_k	The kth component of output data	-			
Z_k	The kth component in the transformed space	-			

ABBREVIATIONS

Symbol	Description	Symbol	Description
ANN	Artificial Neural Network	MCSA	Motor Current Signature Analysis
BLDCM	Brush Less Direct Current Motor	MEC	Magnetic Equivalent Circuit
BPNN	Back Propagation Neural Network	MMF	Magnetic Motive Force
CAE	Contractive Autoencoder	NB	Naïve Bayes
CNN	Convolutional Neural Network	PCA	Principal Component Analysis
CPU	Central Processing Unit	PM	Permanent Magnet
CUDA	Compute Unified Device Architecture	PMSM	Permanent Magnet Synchronous Motor
CWD	Choi-Williams Distribution	PNN	Probabilistic Neural Network
CWT	Continuous Wavelet Transform	REE	Rare-Earth-Element
DAE	Denoising Autoencoder	RNN	Recurrent Neural Network
DT	Decision Tree	ROC	Receiver Operating Characteristic
EDM	Electro Discharge Machining	SAE	Sparse Autoencoder
EMD	Empirical Mode Decomposition	SGD	Stochastic Gradient Descent
EMF	Electromotive Force	SHM	Space Harmonic Method
FDA	Fisher Discriminant Analysis	SIPMSM	Surface Inset PMSM
FEM	Finite Element Method	SLP	Single layer Perceptron
FFT	Fast Fourier Transform	SOM	Self-Organizing Maps
FN	False Negative	SPMSM	Surface Mounted PMSM
FNR	False Negative Rate	STFT	Short-Time Fourier Transform
FOC	Field Oriented Control	SVM	Support Vector Machines
FP	False Positive	TDE	Time Delay Embedding
FPR	False Positive Rate	TMCSA	Transient Motor Current Signature Analysis
FRM	Field Reconstruction Method	TN	True Negative
GD	Gradient Descent	TNR	True Negative Rate
GPU	Graphics Processing Unit	TP	True Positive
GST	Grey System Theory	TPR	True Positive Rate
IMF	Intrinsic Mode Function	VAE	Variational Autoencoder
IPMSM	Interior PMSM	VKF	Vold Kalman Filter
KL	Kullback-Leibler	WGAN	Wasserstein Generative Adversarial Neural Network

KNN	K-Nearest Neighbourhood	WPT	Wavelet Packet Transform
LDA	Linear Discriminant Analysis	WVD	Winger Ville Distribution
LMS	Least Mean Square	ZCP	Zero-Crossing Point
MAP	Maximum A Posteriori	VI	Variational Inference

1 INTRODUCTION

Nowadays, the most challenging issue attracting the researchers' consideration is producing electric energy using limited renewable resources globally. The countries with restricted energy resources have focused their economic and political plans on finding ways to produce electricity using green resources. The main goal of these plans is to produce electricity with high efficiency and protect the environment against the power plants' operation side effects. For instance, the German government decided to turn off most of its nuclear power plants to prevent the hazardous effects of nuclear accidents similar to the Fukushima disaster in 2011 [1]. Therefore, producing electric energy with high efficiency in safe conditions is one of the most important goals of researchers. Electrical machines in both producing and consuming electrical energy have a basic place among all of the electrical equipment.

In 1843, the first electric motor was manufactured by Jacob, which has Permanent Magnets (PMs) in its structure. These magnets has been built from natural magnetite material, making them not perform appropriately in a large-scale machine. Therefore, the manufacturers started using excitation winding instead of PMs in the electric machines' structure. Then induction motors were used extensively in industrial applications because of appropriate features such as high reliability compared to other machines, simplicity in structure and overhaul. However, they are faced with high losses because of using winding in all parts of the machine, which reduces efficiency, power density, and power factors, especially in light load. Besides, the induction machine has a low starting torque and poor speed regulation.

Research on PMs got a new impulse in the 1950s when the ferrites were discovered in nature. Compared to other magnetic materials, such as carbon steel and nickel-cobalt in the 1930s, ferrite had exceptional properties that encouraged the scientists to develop the PM machines using this material. However, the PMs built with ferrite faced some problems in manufacturing because of low magnetic energy. Finally, the Neodymium Iron Boron (Nd-Fe-Br) was discovered, and a revolution happened in the manufacturing of PM machines in the 1960s. It can be noted that these PMs have better properties such as high coercivity, high remnant density, and high magnetic energy compared to other types of magnetic material. Thus, the development of PM machines has continued, and different types of PM machines with different structures have been presented until now.

The importance of using PM machines in wind power plants and wave energy converters for producing electricity from renewable energy has been revealed recently. Using PMs instead of the excitation winding causes to reduce the weight and copper loss of the machine considerably. It can be noted that a PM machine has a high starting torque and a wide performance range of speed which makes it applicable at different speeds and loads. Thus, the power density and power factor of a PM machine have been improved compared to induction machines. Besides, precise control and simple structure are other benefits of this type of machine which is so important in the process of manufacturing and operating. However, these machines have some drawbacks such as higher initial cost than other electric machines because of using expensive PMs. Besides, the controlling of PM machines is performed using a separate expensive digital inverter increasing the cost of these machines. Also, the PM materials are susceptible to temperature and changes of the PM's work point because of operating at high temperature such that demagnetization may become possible. Compared to other types of motors such as induction motors, a reduced weight and a small size with the same power are the most prominent features of PMSM. They are among the best design options for applications where the machine's low volume and low weight are of particular importance.

The torque on the PM motor shaft is generated by the interaction between the magnetic field of the stator winding and the magnetic field of the PMs. Therefore, any fault in one of these two components causes the operation of the motor to be out of the desired state due to changes in the stator and rotor fields. Besides, some important factors such as high level of humidity makes an improper environmental situation for electrical machines. Since some electrical machines are operated in such an environmental condition, and the operating conditions such as load and speed are frequently changing, the possibility of faults in the machine's components has to be taken into account [2]. In electrical machines, faults may occur in any main component, such as the rotor, stator, or other mechanical components. PM machines' faults can be divided into three basic categories: electrical, mechanical, and magnetic faults. Figure 1-1 shows the classification of faults in PM machines. Electrical faults dedicate to faults in the winding of machines. Operating at high temperatures, passing high current from winding, manufacturing errors, and corrosion of winding insulation are some significant reasons for electrical faults. Electrical faults divide into unbalanced resistance, short circuit, and open circuit. As shown in Figure 1-1, the mechanical faults are mainly classified into eccentricity and bearing faults. However, other faults related to the mechanical components of the electric machine, such as unbalanced fault and lamination fault in rotor and stator core, can be related to mechanical faults. Magnetic faults are related to demagnetization in the PM machine. This type of fault is defined only for permanent magnet machines, unlike other faults. Demagnetization disturbs the distribution of the magnetic field in the air gap [3]. In the following, the structure of PM materials is explained and demagnetization fault is investigated in detail.

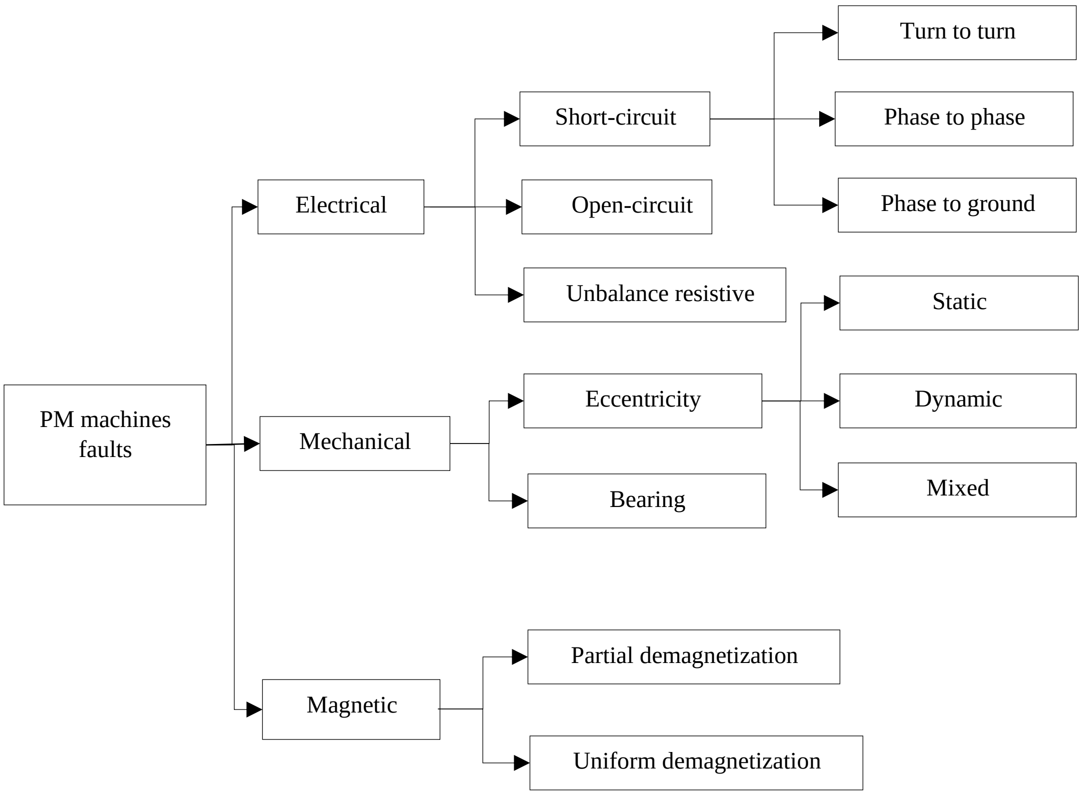


Figure 1-1: Classification of various faults in the PM machines [3]

1.1 Permanent Magnets

This section is dedicated to explain the properties of magnetic materials, especially PM material. According to the electron configuration and reaction of the material to an external applied magnetic field, materials in nature are divided into paramagnetic, diamagnetic, ferromagnetic, and ferrimagnetic materials. When a magnetic field is applied to diamagnetic materials such as Copper ($_{29}Cu$), Silver ($_{47}Ag$), Gold ($_{79}Au$), Silicon ($_{14}Si$), or monocrystalline Graphite ($_{6}C$), the magnetic moment of electrons is oriented in the opposite direction of the external magnetic field, and they start to repulse from the magnetic field.

When paramagnetic materials, such as Aluminium ($_{13}Al$) and Titanium ($_{22}Ti$), are located near to the external magnetic field, the moment of electrons is oriented in a direction similar to the external magnetic field. Therefore, they amplify the external magnetic field slightly and are attracted to it. Besides, when the external magnetic field is removed, they lose their magnetic properties.

In the atomic structure of ferromagnetic materials, unpaired electrons produce magnetic moments. Besides, the magnetic moment tends to align with each other. When an external magnetic field is applied to the ferromagnetic material, their electrons' moment is aligned in the external field's direction and amplifies this magnetic field. In contrast to the material mentioned earlier, when the magnetic field is removed, they keep their magnetic properties. Iron ($_{26}Fe$), Nickel ($_{28}Ni$), and Cobalt ($_{27}Co$), and Rare-Earth-Element (REE) material such as Neodymium ($_{60}Nd$), are some of the ferromagnetic materials.

Ferrimagnetic materials such as Magnetite (Fe3O4) are another type of materials in which the atoms have opposing magnetic moments; however, their size is different, and they cannot neutralize each other. Therefore, the pure magnetic moment remains in the material. When these materials are exposed to an external magnetic field, these residual moments are directed, and the material becomes magnetic. Besides, after removing the external magnetic field, they keep their magnetic properties like ferromagnetic materials.

The PMs are built from ferrimagnetic or ferromagnetic materials, which maintain their magnetic properties after removing the external magnetic fields. PMs built from REE materials such as Nd-Fe-Br and Sm-Co are the best among all types of magnets until now. However, different magnets are produced using ferrimagnetic materials such as Ferrite and Alnico, whose coercivity and magnetic power are lower than REE's PMs [4].

1.2 Demagnetization Phenomenon

When a magnetic material is magnetized, its magnetic parameters can be defined as a linear relation between magnetic field strength (H) and magnetic flux density (B) as follows [5]:

$$B = \mu H \quad , \quad \mu = \mu_0 \mu_r$$
$$\mu_0 = 4\pi \times 10^{-7} \ {}^{\mathrm{H}}/_{\mathrm{m}} \tag{1-1}$$

where μ_0, μ_r are magnetic permeability of free space and relative permeability, respectively. However, in case of cyclic magnetization and demagnetization, it is not possible to calculate the magnetic property based on equation (1-1). According to the magnetic properties of each material, the hysteresis curve can be defined as shown in Figure 1-2. The B-H curve depicts the nonlinear behavior of a ferromagnetic core. As shown in Figure 1-2, the starting point on the curve, point "0", corresponds to an unmagnetized core where both magnetic field intensity and magnetic flux density are zero. As the magnetization current increases in the positive direction, H and B increase linearly until *B* reaches a saturation point at point "A". If the magnetization current decreases to zero, *H* also decreases to zero, but due to magnetic hysteresis in the core, *B* does not reach zero and follows the curve from point "A" to point "B". To reduce *B* to zero at point "B", a demagnetizing force or coercivity is needed. When the magnetic flux density decreases from point "B" to zero, it reaches a knee point where a small change in the magnetic field intensity causes a large change in the magnetic flux

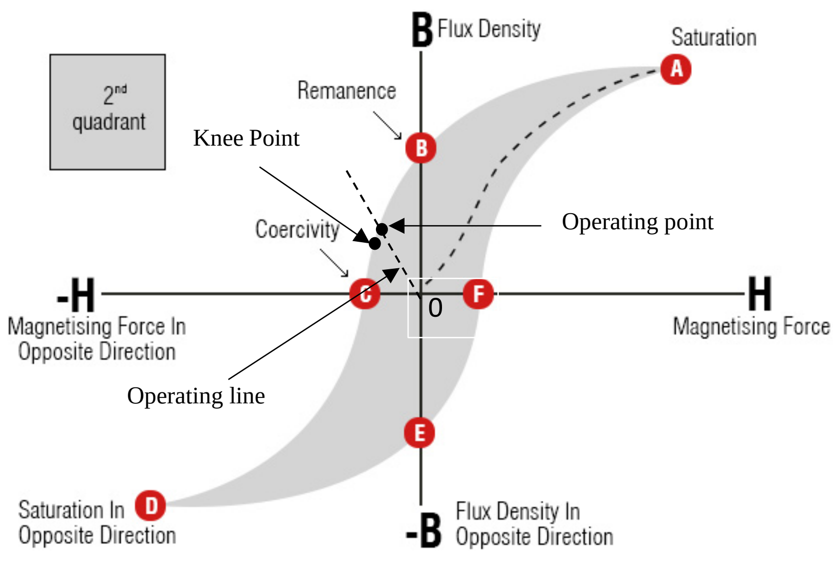


Figure 1-2: Magnetic hysteresis loop or B-H curve [7]

density. This force reverses the magnetic field, causing the molecular magnets in the core to become random again until the core becomes non-magnetic, shown at point "C" on the curve. When the current is reversed, the core is magnetized in the opposite direction, reaching its saturation point in the opposite direction at point "D", which is symmetrical to point "A". If the current decreases to zero, the magnetic hysteresis in the core has the same value as before but reversed, shown at point "E". By reversing the current again in the positive direction, the magnetic flux reaches zero at point "F", and as the current increases further, the core reaches its saturation point at point "A" again. This process repeats in an AC-like periodicity, and the path followed by the B-H curve is known as the magnetic hysteresis loop. This effect shows that the magnetization process of a ferromagnetic core and the magnetic flux density depend on the state of magnetization of the core.

Generally, when a PM is magnetized using an external magnetic field, it doesn’t lose its magnetic properties unless improper condition causes demagnetization. Each type of PM has its temperature’s coefficient, which shows the dependence on temperature changes. For instance, when the temperature coefficient in a PM is large such Nd-Fe-Br, rising temperature increases the possibility of demagnetization. Therefore, the temperature is a significant factor in demagnetization happening in PM machines [6]. As shown in Figure 1-3, increasing temperature (T) causes the displacement of the B-H curve of the PM. Therefore, the intersection of the B-H curve (black line) and operating line (red line), which shows the PM's operating point, is displaced because of increasing temperature.

When the PMs are exposed to high stress, vibrations, or damage, the possibility of changing some of the magnetic moments is high, leading to demagnetization [8]. Besides, oxidation or exposure to an acidic or alkalic environment increases the chance of deterioration of magnetic moments in the PM. However, the probability that demagnetization happens in this condition is low unless over a long time.

The factors mentioned above displace the operating line to lower than the knee point because of the deformation of the B-H curve; however, some other factors cause the PM’s operating line to move. High armature current and external reverse magnetic fields are some factors for demagnetization faults that makes displacement of operating line [9]. The operating point of the magnet can be defined as the intersection of the operating line and the part of the hysteresis curve in the second quadrant called the PM's demagnetization curve as shown in Figure 1-2. For finding the coordinates of the operating line in the B-H curve, Ampère's circuit law can be applied for one section of a PM machine as shown in Figure 1-4, as follows [5]:

$$\oint Hdl = Ni \rightarrow H_{\mathrm{c}}l_{\mathrm{c}} + H_{\mathrm{m}}l_{\mathrm{m}} + H_{\mathrm{g}}l_{\mathrm{g}} = Ni \tag{1-2}$$

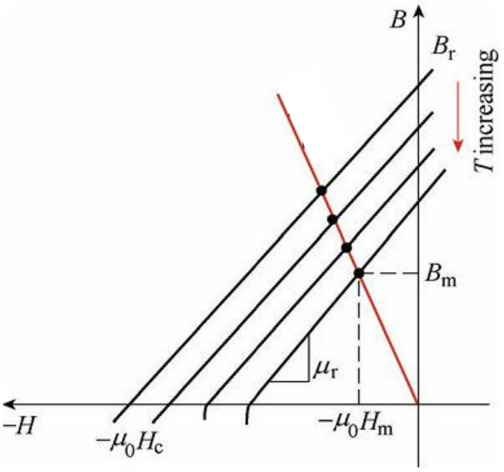


Figure 1-3: The effect of temperature (*T*) on changing the coordination of the B-H curve [10]

where H_{c}, H_{m}, H_{g} are magnetic field strength of core, PM, and air gap, respectively, while N and i indicate the number of winding's turns and current. Besides, l_{c}, l_{m}, l_{g} are the length of flux path in rotor and stator cores, magnets, and air gaps, respectively. They are defined as follows:

$$l_{\mathrm{c}} = l_{\mathrm{cs1}} + l_{\mathrm{cs2}} + l_{\mathrm{cs3}} + l_{\mathrm{cr1}} + l_{\mathrm{cr2}} + l_{\mathrm{cr3}} \tag{1-3}$$

$$l_{\mathrm{m}} = l_{\mathrm{m1}} + l_{\mathrm{m2}} \tag{1-4}$$

$$l_{\mathrm{g}} = l_{\mathrm{g1}} + l_{\mathrm{g2}} \tag{1-5}$$

where $l_{\mathrm{cs1}}, l_{\mathrm{cs2}}, l_{\mathrm{cs3}}, l_{\mathrm{cr1}}, l_{\mathrm{cr2}}, l_{\mathrm{cr3}}$, $l_{\mathrm{m1}}, l_{\mathrm{m2}}$, $l_{\mathrm{g1}}, l_{\mathrm{g2}}$ are the length of the flux path in different sections of stator, rotor, magnets, and air gap specified in Figure 1-4. According to the fact that the permeability of the core is so high [5], it is possible to remove H_{c} from equation (1-2). Therefore, equation (1-2) changes to equation (1-6) as follows:

$$H_{\mathrm{m}}l_{\mathrm{m}} + H_{\mathrm{g}}l_{\mathrm{g}} = Ni \tag{1-6}$$

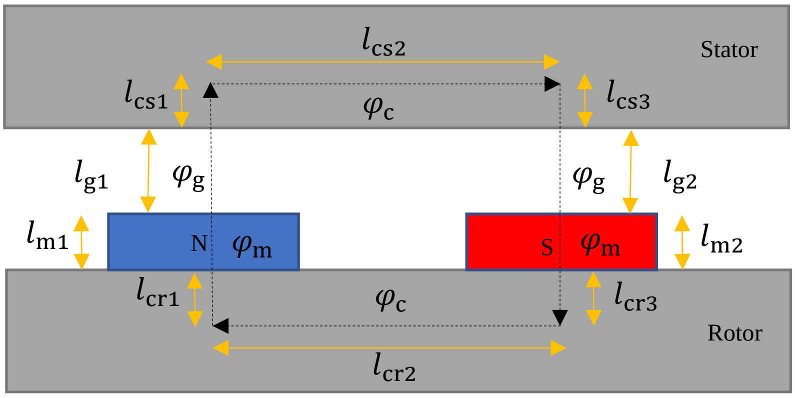


Figure 1-4: One section of PM machines

Besides, all of the PM's flux which goes to the air gap is not effective flux. Therefore, the coefficient k is defined as an effective flux coefficient of PM. The relation between air gap flux φ_g and PM flux φ_m is prepared as follows:

$$\varphi_m = k\,\varphi_g \rightarrow B_m A_m = k\,B_g A_g \tag{1-7}$$

$$H_g = \frac{B_g}{\mu_0} \tag{1-8}$$

where, A_m and A_g are the area of magnet and air gap, respectively. With substituting equations (1-7) and (1-8) in equation (1-6), the coordinates of the operating line can be obtained as follows:

$$\begin{aligned} H_m l_m + \frac{B_m A_m}{k\,\mu_0 A_g} l_g = Ni \rightarrow B_m &= (-H_m l_m + Ni)\left(\frac{k\,\mu_0 A_g}{A_m l_g}\right) \\ &= -\frac{l_m A_g \mu_0 k}{A_m l_g} H_m + \frac{k\,\mu_0 A_g\,N}{A_m l_g} i \end{aligned} \tag{1-9}$$

Therefore, the coordinates of the operating line depend on the size of the PM and the air gap, number of winding's turns and armature current. It can be noted that changing the air gap affects the slope and *y*-intercept of the operating line.

Demagnetization may occur in uniform and partial forms in the machine. Partial demagnetization occurs when one or more PMs are demagnetized, and uniform demagnetization occurs when all of the machine's PMs are affected. When demagnetization occurs in the motor, since the machine wants to continue its normal operation, it is forced to pass more current through its coils to produce the required output power, increasing copper losses and overheating. Therefore, as the machine heats up more, the demagnetization is transferred to other PMs, and the stator current passing through the coil increases. This fault also increases vibration and noise in the machine due to the increase the torque ripples in an electric machine [11]. Besides, the magnetic's flux decreases with demagnetization, and the induced voltage begins to decline. Therefore, timely demagnetization detection prevents the spread of fault to other machine components. Another critical reason for diagnosing fault is the increasing price of PMs in recent years, which shows the importance of accurate detection and replacing healthy PMs with faulty ones.

The chance that the operating line shifts because of factors like changes in the winding's current is higher than the chance that the magnet's B-H curve deforms due to factors such as magnet damage, in critical cases the former may cause a shift in the operating point below the knee point of the B-H curve. Figure 1-5 shows the effect of the current on changing the operating line. As illustrated, when the current increases, the *y*-intercept of the machine's operating line is displaced and causes to displace the intersection point with the B-H curve from above to below the knee point. When the intersection points are displaced and located above the knee point (point "a"), reversible demagnetization happens, and when the intersection is located below the knee point (point "b"), irreversible demagnetization occurs.

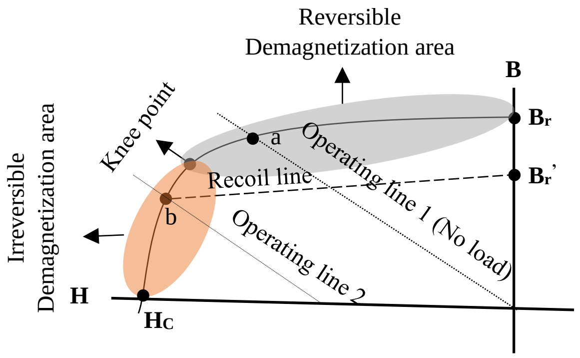


Figure 1-5: Demagnetization curve by changing the working point of the machine

According to the important role of PMs' health in the PMSM's operation, this research work presents a new method for demagnetization detection in the PM machines. This approach is investigated to detect the bearing and eccentricity faults in order to present a comprehensive fault diagnosis method in PM machines. Therefore, the main steps of this research work are prepared as follows:

- Reviewing the prominent research works related to demagnetization detection;
- Prototyping a Permanent Magnet Synchronous Motor (PMSM) for implementing faults in different operating conditions;
- Extracting different signals in the healthy and faulty states for demagnetization, bearing, and eccentricity faults with different severity;
- Proposing a new method for fault diagnosis in PMSM;

1.3 Thesis Organization

This thesis is structured into six main chapters and different subsections. This chapter describes the importance of demagnetization detection in PM machines. First, the importance of producing electric energy using green resources is explained, and then the trend of development of PM machines is discussed. Next, the different types of magnetic materials are described. Besides, the classification of various faults in electrical machines is presented, and the demagnetization phenomenon based on the B-H curve is described in detail. Finally, the foundation of this research work is defined, and the main route of this research work is specified.

In the next chapter, a literature review on different demagnetization detection methods is presented. Firstly, different methods of demagnetization detection are scrutinized then some of the important research is explained. Thesis objectives are presented in the final section of this chapter based on the reviewed research works, and the new method is introduced.

The theoretical basis of the proposed method for diagnosing faults is presented in Chapter 3. Firstly, the machine learning methods are studied. Then the proposed method is investigated and its structure is defined entirely.

Chapter 4 presents some important information about the experimental setup components and simulation of the electric machine. The object of this thesis is not the designing and optimizing of electrical machines; however, it can be helpful to present the mathematical equations of PM machines. Then the machine simulation using Finite Element Analysis (FEA) is presented and compared to the experimental machine for faulty and healthy states. Next, implementing selected faults on the electric motor is explained, and the data acquisition system consisting of sensors and data acquisition cards is investigated in detail.

The proposed method is analysed and developed in Chapter 5. First, the confusion matrix, receiver operating characteristic, and important indexes such as accuracy, precision, recall, and F1-score for evaluating the new method are described. Then, the performance of the proposed method is analysed in rated conditions (rated load and speed) for different separate faults. In addition, obtained results with all data comprising different faults in various loads and speeds are prepared. Next, the effect of implementing the pre-processing methods on the proposed method for separate faults and different combinations of faults and sensors are presented. Finally, the proposed method is compared to some machine learning methods to benchmark this method's performance in the same condition.

The last chapter of this thesis is dedicated to present the conclusion from the analysis of the proposed method. According to the obtained results, some suggestions are prepared to continue this research work in future as well.

2 STATE OF THE ART

The main goal of this chapter is to review the state of the art of research on demagnetization detection. The structure of this chapter is divided into two main sections. First, the demagnetization detection methods are described in two subsections; model-based and signal-based. Then the main objectives of this research work based on the selected method are presented.

2.1 Demagnetization Detection Methods

Fault detection methods are generally divided into two main categories, based on models and signals. Techniques that fall into the signal-based category are more widely used in industry, while model-based methods have a faster simulation time if the correct model of machine modelling is available.

2.1.1 Model-Based Detection Methods

Model-based fault detection methods estimate the machine parameters based on the mathematical model of the machine. The residual, the difference between the estimated model and the mathematical model, is generated according to the measured input and output signals in the model-based approaches. Then, the states of the system are estimated based on the detection method. Finally, the fault is detected based on the difference between the states in normal and faulty states. In the model-based detection methods, the correct model of the machine is required. In general, the mathematical model of a PM machine can be expressed as follows [12]:

$$\begin{cases} u_{\mathrm{d}} = R_s i_{\mathrm{d}} + L_{\mathrm{d}} \dfrac{di_{\mathrm{d}}}{dt} - \omega L_{\mathrm{q}} i_{\mathrm{q}} \\ u_{\mathrm{q}} = R_{\mathrm{s}} i_{\mathrm{q}} + L_{\mathrm{q}} \dfrac{di_{\mathrm{q}}}{dt} + \omega L_{\mathrm{d}} i_{\mathrm{d}} + \omega \lambda_{\mathrm{pm}} \end{cases} \tag{2-1}$$

$$T_{\mathrm{e}} = \frac{3}{2} p \left[\lambda_{\mathrm{pm}} + \left(L_{\mathrm{d}} - L_{\mathrm{q}}\right) i_{\mathrm{d}}\right] \tag{2-2}$$

Where u_{d} ,u_{q} ,i_{d} ,i_{q} represent the voltages and currents in the direct (d) and quadrature (q) axis, respectively. Also, R_{s} is the stator winding resistance, p is the number of pole pairs, L_{d}, L_{q} are the inductance of the d and q axes, λ_{pm} is the constant linkage flux of PM, ω is the angular velocity, and T_e is the electromagnetic torque of the machine. In

[12], a model is used to estimate the linkage flux of the rotor using the Kalman filter based on the above equations. In this method, the distortions of PM flux are investigated using an analytical method, and the demagnetization in the machine is detected.

Numerical simulation using Finite Element Methods (FEM) is one of the PM machines' most effective fault analysis methods. In FEM, the magnetic fields are calculated based on different factors such as the magnetic circuit, the winding distribution, the slots' location, and the nonlinearity behaviour of the ferromagnetic material. In [13], an analytical model is proposed for brushless permanent magnet motors. This method follows a well-known way for solving the vector potential equation for the magnetic field around a permanent magnet with a certain magnetization in a simplified 2-dimensional setting. Under these particular circumstances, the vector potential equation takes the form of a Laplace equation in the air region (in spherical coordinates) and a form similar to a Poisson equation in the magnetized material. Then the model is verified by finite-element software, and the Electromotive Force (EMF) signal is extracted. Next, the difference between the extracted EMF signal and corresponding data from a motor to be examined (which is in [13] for simplicity replaced by simulated FEM data) is entered into a gradient algorithm to find the location of the fault. In this way, the algorithm can identify the point of maximum change, which is typically associated with the location of the fault. In this research work, partial demagnetization is modelled by replacing several areas with air or reducing the residual flux density. However, this method has the problem of simulating using 2-D FEM to find the fault location, and using 3-D FEM is needed. Thus, an increasing time consumption of this method can be expected in some cases.

In [14], an algorithm based on the machine's structure is presented in which the inductance changes due to the changes of the linkage flux are calculated, and the error is specified using the least-squares method. Therefore, demagnetization is modelled according to changes in the flux linkages of PMs. In [15], an analytical model of an axial-flux PM synchronous motor is proposed to detect eccentricity and demagnetization faults in the various loading conditions. This model is evaluated with practical results, and it is found that this model is suitable for all axial-flux machines with two air gaps, both distributed and concentrated, and with any topology and structure without the use of additional sensors. As this method only requires few calculations, it can also be used to detect faults online.

The simultaneous use of FEM and magnetic field analysis requires a high computational time; however, they are in a proper position among modelling methods in terms of accuracy. The analytical technique applying the theory of electric machines and Maxwell equations is computationally fast. As these methods do not require analysing all machine components, many simplifying hypotheses, such as omitting the effects of slots and saturation, are admissible for calculations. In order to mitigate the high time consumption inherent in the FEM, a supplementary technique known as the FEM auxiliary method is introduced. This method involves the calculation of modeling parameters such as d-q parameters using various techniques including FEM, finite element models based on phase variables and resistance network, and Field Reconstruction Method (FRM) [16]. These methods have a reasonable accuracy, and computational time. Therefore, they can be used in real applications. In methods based on resistive networks and FRM, where a slot is simulated, the computational time is better than FEM. However, these methods require a high computational time compared to data-based methods. In [17], a valuable analytical model for an axial flux synchronous motor is proposed. In this method, scalar potential equations are calculated for the different sections of magnets. This model estimates the magnetic coefficients of the machines based on a three-phase current, the electromagnetic properties and geometry of the machine as inputs, the EMF, and the terminal voltage as outputs. According to the square waveform of magnetization which is corrected using the magnetization coefficient for each magnet, demagnetization is modelled. Therefore, demagnetization is modelled according to the change in these coefficients. This model is evaluated by experiments and FEM simulations.

In [18], two-dimensional FEM is first applied to a synchronous PM motor. Then the d-q inductances related to the linkage flux and torque fluctuations are estimated and used in the dynamic d-q dynamic model of the PM machine. Consequently, simplifying assumptions with estimated values can reduce simulation time. In [19], an analytical model based on the permeance network and superposition theorem in an axial flux permanent magnet motor is proposed. In this model, the air gap potential vector is calculated using FEM, and the effect of saturation and temperature changes on machine performance is investigated. The extracted model is validated using experimental data. In [20], a resistive model is proposed to investigate the partial demagnetization in-wheel surface permanent magnet motor. The nonlinear behaviour of the core is considered for analysing the slot effect. This method calculates the reluctance of the air gap network for any rotor position, and the torque and back-EMF are calculated for various levels of saturation. According to the extracted results, it is found that this method is a suitable option for optimizing geometry.

In [21], FRM is used to predict torque ripples, sound noises, and magnetic fluxes of stator teeth. In this method, the magnetic field generated by the stator used a slot in each phase with a PM field to determine the distribution of the magnetic field and the electromagnetic force. This method calculates the tangential and normal magnetic flux density components using two-dimensional and three-dimensional magnetostatics analysis. Other machine signals are obtained using these components and different analytical formulas. Also, there are analyses based on system identification methods [22]. In these methods, the time-series data such as voltage and current are collected to train the model’s parameters. Then the model is defined for the system in a way that is compatible with the existing system. Also, the structure and selected parameters of the model are adjusted to minimize the error.

Table 2-1: Evaluating demagnetization modeling methods

Method	Accuracy	Memory	Time-consumption
FEM [13-15]	Very high	Very high	Very high
MEC [18-20]	High	Low	Medium
Maxwell equation [13]	Medium	Low	Low
System identification [22]	High	High	Medium
FRM [21]	High	Medium	Low
Estimating models’ parameters [12]	Medium	Low	Low

Table 2-1 compares different model-based methods described in this section. As shown in this table, the aforementioned main methods based on various indexes consisting of accuracy, the memory needed for calculation, and time-consumption indexes are evaluated. According to this analysis, each described method has some advantages and disadvantages. It is essential to select a proper method based on the machine geometry and software infrastructure for detecting demagnetization. In the continuation of this research, FEM is used for simulating the prototyped PM motor in faulty and normal conditions. The goal of this investigation is to determine the correlation between the results obtained through simulation and those observed experimentally.

2.1.2 Signal-Based Detection Methods

In this approach, the measured signals are used to detect the fault, and the machine's model is not required. According to the signal-based detection scheme shown in Figure 2-1, the signals are collected using the acquisition instruments. Then, the important features of the measured signal are extracted in the process and based on a specific criterion, and the decision is made to detect the fault. In signal processing, two important categories of signals are defined based on data distribution in the space: stationary and non-stationary signals. Statistical data characteristics (mean and variance) remain constant in stationary signals though they change for non-stationary situations.

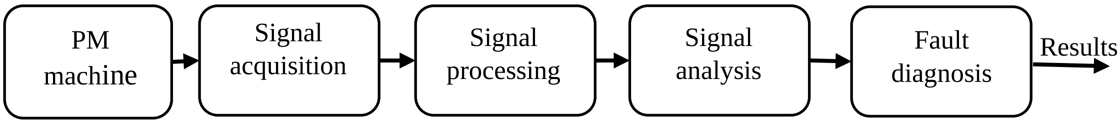


Figure 2-1: Signal-based detection schematic

Before reviewing demagnetization detection methods based on different types of signals, finding the related frequency components relevant for demagnetization can be helpful according to the study in [23]. In this analysis, the induced voltage of one slot of a PMSM is computed. It is assumed that the induced voltage has a sinusoidal shape, and the voltage drop is considered a multiple of a rectangular pulse in the main component of the sine wave. Therefore, the Fourier series of a rectangular pulse is presented as follows [23]:

$$\mathrm{rect}(t) = \frac{1}{2p} + \sum_{k=1}^{\infty} \frac{2}{k\pi} \sin(\pi k d) \cos\left(\frac{2k\pi f_e t}{p}\right) \tag{2-3}$$

In this regard, p and f_e are related to the number of pole pairs and the main frequency component, respectively. This waveform has a duty cycle of $d = \frac{1}{2p}$. Therefore, the reduced voltage due to multiplying of square wave and the sinusoidal wave is presented as follows:

$$y(t) = \cos(2pf_e t) \left[\frac{1}{2p} + \sum_{k=1}^{\infty} \frac{2}{k\pi} \sin(\pi k d) \cos\left(\frac{2k\pi f_e t}{p}\right)\right] \tag{2-4}$$

By simplification, the formation of equation (2-4) changes to:

$$y(t) = \frac{1}{2p}\cos(2pf_\mathrm{e}t) + \sum_{k=1}^{\infty}\frac{1}{k\pi}\sin(\pi kd)\cos\left(2k\pi f_e t\left(1 \pm \frac{k}{p}\right)\right) \tag{2-5}$$

The induced voltage in a slot of a phase is [23]:

$$\frac{\mathrm{d}\lambda_\mathrm{slot}}{\mathrm{dt}} = V_\mathrm{slot}\cos(2\pi f_e t) - V_\mathrm{slot}K_\mathrm{dem}y(t) \tag{2-6}$$

In this regard, λ_slot, V_slot and K_dem are related to the flux linkage, the induction voltage in the slot, and the severity of demagnetization, respectively. Also, by replacing (2-5) in (2-6):

$$e_\mathrm{slot,1} = \frac{\mathrm{d}\lambda_\mathrm{slot}}{\mathrm{dt}} = V_\mathrm{slot}\left(1 - \frac{K_\mathrm{dem}}{2p}\right)\cos(2\pi f_e t) - V_\mathrm{slot}K_\mathrm{dem}\sum_{k=1}^{\infty}\frac{1}{k\pi}\sin(\pi kd)\cos\left(2k\pi f_e t\left(1 \pm \frac{k}{p}\right)\right) \tag{2-7}$$

As it is clear from equation (2-7), the main component of induction voltage is reduced by a factor of $(1 - \frac{K_\mathrm{dem}}{2p})$ and the harmonic frequencies due to demagnetization have the value of equation (2-8).

$$f_\mathrm{dem} = \left(1 \pm \frac{k}{p}\right)f_\mathrm{s} \quad k = 1,2,3, \ldots. \tag{2-8}$$

Also, the amplitude of each harmonic is:

$$A_\mathrm{dem} = \frac{V_\mathrm{slot}K_\mathrm{dem}}{k\pi}\sin\left(\frac{\pi k}{2p}\right) \tag{2-9}$$

Considering the location of adjacent slots (180 degrees electric) and the direction of the current, the induced back-EMF in slot can be calculated as follows [23]:

$$e_\mathrm{slot} = \frac{d\lambda_\mathrm{slot}}{dt} = (-1)^{m-1}V_\mathrm{slot}\left(1 - \frac{K_\mathrm{dem}}{2p}\right)\cos\big(2\pi f_e t - \pi(m-1)\big) - (-1)^{m-1}V_\mathrm{slot}K_\mathrm{dem}\sum_{k=1}^{\infty}\frac{1}{k\pi}\sin\left(\frac{\pi k}{2p}\right)\cos\left(2k\pi f_e t\left(1 \pm \frac{k}{p}\right) - \pi(m-1)\left(1 \pm \frac{k}{p}\right)\right) \tag{2-10}$$

In this regard, the value of m represents the number of harmonics, and also the direction of current is considered in the coefficient $(-1)^{m-1}$. By integrating from relation (2-10), the linkage flux of one slot is calculated as follows:

$$\lambda_{\mathrm{slot}}(\mathrm{t}) = \int e_{\mathrm{slot}}(\mathrm{t})\,dt = (1)^{m-1} N\lambda_1 \left(1 - \frac{K_{\mathrm{dem}}}{2p}\right) \sin\big(2\pi f_e t - \pi(m-1)\big) - (-1)^{m-1} N\lambda_1 K_{\mathrm{dem}} \sum_{k=1}^{\infty} \frac{p}{k\pi(p \pm k)} \sin\left(\frac{\pi k}{2p}\right) \sin\left(2k\pi f_e t \left(1 \pm \frac{k}{p}\right) - \pi(m-1)\left(1 \pm \frac{k}{p}\right)\right) \quad (2\text{-}11)$$

where $N\lambda_1 = \frac{V_{\mathrm{slot}}}{2\pi f_e t}$. Therefore, the total induced voltage for all slots is equal to the induced voltage of a phase, considering the type of winding, the number of poles, and the number of slots. The effects of the number of poles and slots are studied in [24] using MEC. According to this valuable study, if the coil distribution is symmetrical, some of the calculated harmonics components by equation (2-8) are removed. In other words, when the ratio of slots to poles number has a value multiplied by 3/2, fractional-order harmonics are not active. A table in [25] shows the active frequencies in the different slot and pole combinations. In general, the phase angle of the fault harmonics depends on the location, the number of stator phase coils, and the number of rotor poles. Therefore, detecting demagnetization in all PMSMs using faulty harmonics reported in equation (2-8) is not possible.

Non-stationary signals in electric machines are extracted from the machine’s terminal that does not operate under stable conditions. There are no stationary working conditions in the industry, and the working point of the machine is constantly changing in different states of speed and load. In this situation, the condition monitoring of the machine faces various challenges. Indeed, the detection methods for stationary signals cannot be applied for non-stationary signals. The prominent signal-based methods are presented in the following.

2.1.2.1 Demagnetization Detection Using the Current Signal

The current signal is one of the best signals obtained from the machine terminals because it can continuously show the essential features of the machine at different loads and speeds. Besides, it can be constantly collected without disassembling the electrical machine's components.

The Motor Current Signature Analysis (MCSA) method, which is widely used in industry, is of great importance [30]. The basic idea of this method is that any fault and asymmetry in electric machines are reflected in the current signal. Therefore, the machine's condition is monitored with the correct interpretation using the proposed method. MCSA online monitoring does not require additional sensors. This method in demagnetization detection has an improper ability due to the fault symmetry, in which any symptom of fault cannot be seen in the frequency spectrum. Also, changes in load torque and static eccentricity may generate similar faulty harmonics, which poses many detection problems. Besides, the pattern of fractional order harmonics in the induced voltage and current for complete demagnetization is not found in a synchronous motor with a concentrated winding. Given the points made about the MCSA method, it does not have high reliability for fault detection, and it needs further study.

Demagnetization detection using the zero component of the current signal in star-connected PM machines is proposed in [26]. First, the stator and the zero component of the current signal in the rated load are taken for all operating speeds, and the Fast Fourier Transform (FFT) is applied. Then the results are analyzed to detect demagnetization. The main problem of Fourier transforms in fault detection is the disability of detecting the faults in non-stationary signals properly. Therefore, time-frequency analysis methods are proposed to apply to non-stationary signals.

The Transient Motor Current Signature Analysis (TMCSA), which is performed using advanced signal methods, is one of the most widely used methods for fault detection using dynamic signals [27]. The different approaches that have been applied to this signal include short-term Fourier transform (STFT), wavelet transform, bidirectional transform, and Hilbert transform. STFT is proposed as an extended Fourier transform method for dynamic signals. Determining the shape and size of the window according to the type of problem is one of the most significant challenges for this method.

To solve this problem, multi-resolution signal methods such as wavelet transform are used. Both continuous and discrete wavelet transform methods are used to detect demagnetization [6,18]. Continuous wavelet transform provides the time-frequency evolution of the signal. While in the discrete wavelet transform using sequential high-pass and low-pass filters at appropriate frequency ranges, the fault components are extracted. This is why continuous wavelet transform calculations are longer than discrete. In contrast, CWT is more suitable for faults such as slow demagnetization. In [28], a valuable research work in demagnetization detection using discrete wavelet transform is presented. This research work consists of four parts. First, a digital adaptive filter is used to remove the main component of the signal. Then, by applying a discrete wavelet transform to the remaining signals, their features are extracted, and the dimension reduction is performed using the linear discriminant analysis method. Finally, support vector machines (SVMs) are used to classify and diagnose the type of faults. However, wavelet transform can solve the problem of frequency and time resolution, like STFT, this method also needs to determine the type of mother wavelet according to the problem. Also, because demagnetization causes small frequency components, using a sharper mother wavelet cannot be a proper option for fault detection.

Unlike linear methods in which the signal is decomposed into primary components, the quadratic dependent methods study energy distribution over time and frequency. The general form of these distributions is known as the Cohen class, which is determined using the kernel function, and the distribution is harvested with different properties. One of this category's most widely used members is Winger-Ville Distribution (WVD). Since this method can calculate the signal energy at any interval, this distribution has high accuracy. The main problem of these methods is the interaction in energy distribution that causes information duplication and confusion in the distribution. Another problem is the amount and shape of the distribution that depends entirely on the kernel type [29]. The Choi-Williams Distribution (CWD) is another member of the Cohen class, in which it is possible to reduce the interaction by adjusting the parameters of the kernel function. Similarly, Zhao-Atlas-Marx and Born-Jordan are other members which have been presented in various applications due to their ability to eliminate interaction and accuracy [30].

In [35], Hilbert transform is presented for entirely sinusoidal signals with a zero-reference level and is unsuitable for non-stationary signals. To solve this problem, an experimental method called Empirical Mode Decomposition (EMD) allows Hilbert's application through the screening process by producing completely oscillating functions called Intrinsic Mode Functions (IMFs). This decomposition is similar to the Fourier transform, except it works in time and frequency domains. Hilbert Transform makes it possible to focus on appropriate IMFs for fault detection.

The Hilbert transform detects demagnetization in a synchronous motor under variable speeds. According to the experimental work in [31], the health and demagnetization with 50% severity are separated using different IMFs. This decomposition is similar to wavelet decomposition, in which the signal is decomposed into high and low-frequency components, and one of the high or low-frequency coefficients with the proper level is used for fault detection. This method does not require basic information similar to the Hilbert transform, which is updated internally.

The EMD algorithm faces the end effect, which generates improper components and over-estimating the signal's energy [32]. EMD has been developed in recent years to solve these problems, and a new method is proposed based on pre-processing of the fault using CWD in [33]. In this method, the specificity of the signals is obtained using the box-counting method. This method determines the complexity of the nonlinear signal waveform. The Hilbert method has an excellent computational complexity compared to other methods, and also the discrete wavelet transform method has good clarity and accuracy. It should be noted that when dealing with signals that the fault is progressing rapidly, methods with minor calculations are needed; however, demagnetization is one of the slow faults. Therefore, the issue of computational complexity is not very important. In [34], using Vold-Kalman-Filter (VKF) to detect demagnetization is another effective method for finding the fault severity on different load and speed conditions. In this method, the main component of the fault is detected and used as an indicator in fault detection. Although time-frequency methods follow frequency harmonics correctly, most consider a specific frequency spectrum at different time intervals. Whereas in harmonic trackers using the Kalman filter, the desired frequency harmonic is detected by a filter whose centre is on the selected frequency. Therefore, the related harmonics to demagnetization fault are tracked, and other harmonics are removed as noise. It is clear that some important information is lost, and there is no chance of detecting other faults. In other words, when the condition monitoring of an electrical machine with different faults is considered, it is not a proper way to use this method. In [35], an analytical method is proposed using the d-q form to detect and diagnose demagnetization, and inter-turn short-circuit faults in a PM synchronous motor. An important point to note is that demagnetization and inter-turn short-circuit faults are two completely different faults which have the same symptoms in output signals. A new index based on the current angle is proposed in this method. Increasing and decreasing of the PM degree and inter-turn short circuit fault causes a change in this angle. Therefore, it can be used as an indicator to indicate demagnetization and inter-turn short-circuit faults.

In recent years, deep neural network methods have been developed as new ways to diagnose faults, and a lot of beneficial research work has been presented in fault detection. In [36], wavelet packet transforms (WPT), and one-dimensional Convolutional Neural networks (CNN-1D) are used to extract the current signal characteristics. Both of them have good accuracy using this method. This research work reports 98.1% and 98.8% for WPT and CNN-1D, respectively. First, the data are decomposed using wavelet transform to detail and approximation coefficients. Then the FFT is implemented on the detail coefficients in different decomposition levels. Therefore, the data are prepared to enter the convolutional neural network. Because of using fewer practical results in this research work, the system was not over-trained, and the performance was acceptable. However, the drawback of this research work can be considered as the lower number of faults scenarios and the lack of various faults' severity.

In [37], research work shows detecting demagnetization using autoencoder and k-means clustering. In this research, current, voltage, speed, power, and torque signals are extracted from the motor driver with a 1kHz frequency rate for three classes of data consisting of normal, demagnetization with 10% severity and 25% severity. According to the presented results, the classification accuracy of this method is around 96%. This method is investigated at different speeds and loads. However, the number of classes for having a proper view about the performance of proposed method is not enough.

In [38], innovative research work is presented on the combination of Variational Autoencoder (VAE) and Wasserstein Generative Adversarial Neural Network (WGAN) to extract important features of data when enough faulty cases are not in access to propose a comprehensive method for detecting faults. First, VAE-WGAN is proposed as a feature extraction method. Then the mean and variance of data are used as an index with two features to fit the actual data. Finally, using this index, the classification is done. This method has some drawbacks, such as not investigating the overfitting problem because of the high number of iterations. Besides, the number of samples in the experimental work is not enough to evaluate the network correctly. In this research work, A-phase current, B-phase current, electromagnetic torque, and negative sequence current are used to analyse this method; however, the negative sequence current and electromagnetic torque extract two-dimensional features. According to the obtained results from this method, it is found that VAE has an extraordinary capacity for extracting important features of input data in a bigger size. Therefore, it can be a good idea to enter different types of data, such as vibration data, to investigate the performance of VAE for other kinds of data.

According to the various proposed methods for the current signal, faults are detected online without needing to disassemble the machine components. Different faults such as demagnetization, eccentricity, and turn-to-turn short-circuit faults have similar frequency components in the current signal. Therefore, using methods such as FFT, which investigates the generated harmonics of the machine for detecting faults, is not completely useful. Besides, the accuracy of these methods is affected by load changes, converter and controller harmonics, and different loading conditions. Also, the complexity of applying different time-frequency methods such as STFT and wavelet transform is high. It is needed to define various parameters such as length and width of windows size in STFT, selecting the proper mother wavelets and proper decomposition levels in Wavelet Transform based on the machine's condition and faults.

The methods such as WVD, which calculates the energy of data in different time intervals, are other methods in which finding the proper kernel function increases the complexity of the technique. Besides, duplication of information is another significant drawback of these methods. HHT and VKF are proper methods for detecting faults; however, they are not applicable for distinguishing between demagnetization and other faults. Demagnetization detection based on the deep learning methods has valuable research results, encouraging the researchers to continue working on this route. However, these research works have some drawbacks, such as insufficient data for analysis. According to the high capability of this method to analyse big data, these methods should be analysed using different faults in different states of operation.

2.1.2.2 Demagnetization Detection Using the Voltage Signal

Generally, measurement of the induced voltage of an electric machine is not extracted directly unless it is measured indirectly using techniques such as the Space Harmonic Method (SHM) [39]. However, the sensitivity of the extracted indicators to machine parameters and temperature changing reduces the accuracy and reliability of these methods. In addition, fractional-order harmonics cannot be used to detect the complete demagnetization in the concentrated winding. However, the zero-sequence of the third spatial component of the induced voltage can be calculated from the zero point of the star connection [39]. Also, to eliminate the converter and regulator effects on the current signal, the detection is checked using the zero component of the voltage. In this method, a balanced three-phase resistance network is required to separate the zero component of the voltage from the converter's effect. In fact, according to Figure 2-2, the zero component of voltage is the difference between the neutral point's voltage of the stator's windings and the virtual ground. As shown in Figure 2-2, the zero-sequence voltage can be defined as follows [10]:

$$\begin{cases} V_0 = \frac{1}{3}(V_\mathrm{a} + V_\mathrm{b} + V_\mathrm{c}) - \frac{d\lambda_{\mathrm{pm,0}}}{dt} \\ V_0 = V_{0,\mathrm{c}} + V_{0,\mathrm{m}} \\ \lambda_{\mathrm{pm,0}} = \sum_{v=3n.n=1,3,5,\ldots} \lambda_{\mathrm{pm},v} \cos(v\theta - \theta_v) \end{cases} \tag{2-12}$$

$$\begin{cases} V_{0,\mathrm{m}} = -\frac{d\lambda_{\mathrm{pm,0}}}{dt} \\ V_{0,\mathrm{C}} = \frac{1}{3}(V_\mathrm{a} + V_\mathrm{b} + V_\mathrm{c}) \end{cases} \tag{2-13}$$

where $\lambda_{\mathrm{pm,0}}$ is the zero component of the linkage flux, $\lambda_{\mathrm{pm},v}$ is the v^{th} harmonic amplitude of the linkage flux, θ is the electrical angle, θ_v represents the electrical angle that corresponds to the position of the stator winding where the higher harmonic flux linkage ($\lambda_{\mathrm{pm},v}$) is generated, and V_a , V_b , V_c are the phase voltages, respectively. As it is entirely straightforward, the zero component of the voltage from the converter ($V_{0,\mathrm{C}}$) has no effect, and only the third-order harmonic and its coefficients affect the calculation. Therefore, to detect demagnetization:

$$V_{0,\mathrm{m}} = -k\frac{d\lambda_{\mathrm{pm,0}}}{dt} \tag{2-14}$$

where k is called the degree of demagnetization, $k = 1$ is for the healthy state, and the lower value is for more demagnetization on the machine. The detection is performed using the zero-voltage component. The major drawback of this method is the unavailability of the neutral point of some machines. In [40], an applicable method is proposed using Zero-Crossing Points (ZCP) of induction voltage in a BLDC motor. This study shows that demagnetization causes many changes in ZCP points. This transient state is much more noticeable than the induced voltage drops in the machine. This detection method is suitable for BLDC motors controlled without sensors and using zero-crossing points.

In [41], a method investigates the direction of voltage changing in the d and q axis on the d-q plane for a PM synchronous motor. The static eccentricity, demagnetization, and turn-to-turn fault are examined. According to the extracted results, the voltage decreases in the d-axes and increases in the q-axes for eccentricity. Also, both d and q components are reduced for demagnetization and increased for turn-to-turn fault. The most important advantage of this method is not using other equipment to detect the fault. However, the saturation of the machine at a high load prevents the correct detection of eccentricity fault.

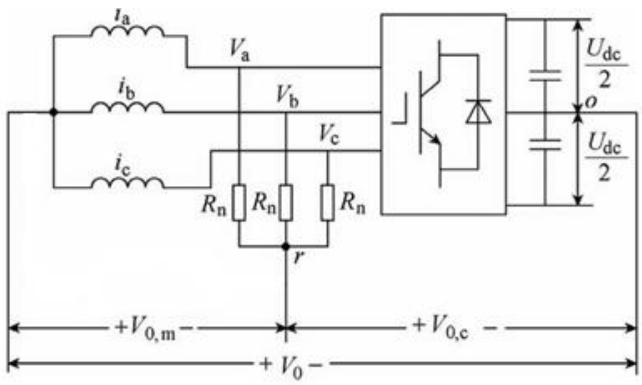


Figure 2-2: A schematic of how to connect a PM machine to an inverter to form a virtual neutral point [10]

In [42], the static eccentricity and the turn-to-turn faults are investigated. First, the voltage and current phase signals are collected. Next, the FFT is applied to collected data to extract important frequency components. Finally, LDA is implemented to classify and estimate different faults with different severity. However, this method in finding of faults' location is not completely successful.

As presented in this section, different methods are proposed to detect demagnetization using voltage signals. Some methods are based on measuring the Back-EMF signals indirectly, and some other is based on measuring the terminal voltage of the machine. As described, measuring the induced voltage is not possible in all the machines; however, it is possible to calculate using different methods such as SHM, ZCP, or analytical methods. Also, these methods have some drawbacks, such as inverter and load effects and not proper ability to find the severity and location of faults.

2.1.2.3 Demagnetization Detection Using the Vibration and Noise Signals

When a fault occurs in a PM machine, the magnetic field of the air gap is disturbed. As a result of asymmetric induced voltage, torque fluctuations, speed changes, vibration, and noise appear at the machine's output. Therefore, vibration and sound signals can be used to detect demagnetization. Partial demagnetization plays an essential role in producing low vibrational modes that produce unbalanced forces. Therefore, detecting faults is possible according to the increasing or decreasing of the produced noises with different spectral patterns compared to other machines' faults. In [43], detection is performed using vibrating acceleration and frequency analysis. In this research work, the double multiple of the main frequency of the vibration signal introduces as an important indicator. In the demagnetization state, the important acceleration components are found in the low frequency, which means there is an unbalanced magnetic pull in this condition. Also, the inter-turn short-circuit fault causes an increase in the circulating current in the winding, distorting the magnetic flux. Finally, it generates different harmonics in a wide frequency range.

In [44], the vibration characteristic proportional to partial demagnetization in the vector-controlled synchronous motor is investigated using three-dimensional analysis and experimental results. This study shows that the amplitude of some vibration components is approximately proportional to the level of demagnetization and load torque. In addition, vibration and torque are beneficial for approximating the demagnetization level of a PM under vector control. However, this method needs many measurements to investigate the demagnetization level in each case. Sound analysis can be a complementary method in the presence of noise and severe disturbances. Model analysis can also be used to investigate the occurrence of the resonance phenomenon and the noisiest faults.

In [45], a valuable work is done on deep neural network using Recurrent Neural Network (RNN) based variational autoencoder (VAE) for classifying 15 classes of data consisting of normal, unbalanced, loosening, and mixing faults in different severity. In this method the vibration data is collected from three accelerometers in three dimensions. The results show that this method gets the 98% accuracy for detection for a large number of data. Besides, the performance of this method is enhanced when linear discernment analysis and principle component analysis are used as the pre-processing methods. This research work shows that using variational autoencoder has an important effect on reducing dimensionality for classifying data. However, the number of scenarios in this research work is low and proper evaluation from this research cannot be achieved. Therefore, it can be a good idea to evaluate this method using high number of scenarios for different faults and different signals such as current.

2.1.2.4 Demagnetization Detection Using the Magnetic Flux Signal

Demagnetization disrupts the distribution of magnetic flux in PM machines. Therefore, checking the symmetry of the reduced magnetic flux in the machine is valuable in fault detection. The distribution and severity of magnetic fields are measured by Gauss meter and Hall effect sensors; however, these methods are considered invasive methods with machine disassembly. Besides, these methods cannot be applied to some machines with different structures, such as interior PM motors. Also, in [46], a low-cost Hall effect sensor in a PM synchronous motor is used for detection, and it is found that placing the sensor in the radial direction increases the sensitivity of the detection method. Therefore, the methods for estimating magnetic flux using the machine models are proposed. In these methods, signals such as flux linkage, torque in BLDC motors, magnetic factors, and magnetic flux in the tooth are estimated [47]. In this model, the measured voltage and current are used as input to the model. In a BLDC motor, the estimated induced voltage is calculated as follows:

$$\hat{E} = V - 2r_{\mathrm{s}}I_{\mathrm{dc}} \tag{2-15}$$

In this equation, the value of V is the phase voltage, r_{s} is the stator phase resistance, and I_{dc} is the dc-link current. By dividing the ratio (2-15) by the rotor speed (ω_{r}), the average torque constant is calculated according to equation (2-16).

$$\widehat{K_t} = \frac{V - 2r_{\mathrm{s}}I_{\mathrm{dc}}}{\omega_r} \tag{2-16}$$

This value is a proper criterion for measuring the quality of a PM. This method detects partial and complete demagnetization faults without being affected by other faults. The dependence of machine parameters such as stator resistance and inductance on temperature and saturation causes uncertainty in the final results. In addition, this method is incapable of detecting minor faults.

In [48], the location and severity of demagnetization are detected using a residual definition by placing a coil and using an analytical method for the induced voltage. In the unsaturated conditions, the probe coil measures the sum of the magnetic flux of the PM machine and the generated flux of armature. To have a proper fault detection method with high reliability, the inverter's noise is eliminated. Also, the main component of the winding's voltage and the forming components of two fluxes are separated.

In [29], the magnetic field is measured using twelve search coils in the healthy and faulty states in a PM machine with different degrees of demagnetization as shown in Figure 2-3. The reduced magnetic flux in each part of the machines shows the occurrence of demagnetization. In this method, load and transient mode do not affect fault detection, and there is no need to use pattern recognition methods. The severity of demagnetization can also be determined using this method. In addition, the proposed method has an excellent ability to diagnose demagnetization, eccentricity, and turn-to-turn short-circuit faults. The main problem of using flux sensors is the invasiveness and installation of the measuring equipment during the construction and winding of the machine. However, using the flux-gate sensor installed on the machine frame to measure the magnetic flux solves this problem to some extent. Besides, some machines that use hall effect sensors in their structure don't need to use another sensor to detect the fault.

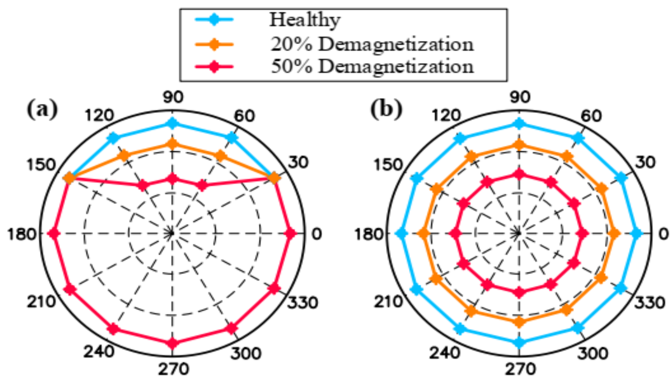


Figure 2-3: The magnetic field under partial and complete demagnetization conditions [29]

In [49], a new algorithm is proposed to detect demagnetization and bearing faults in a BLDC motor. This method uses Hall effect sensors for detecting fault which is used to control speed in BLDC motor. Therefore, no accelerometer or current sensor is used to detect the faults. In this method, a combination of continuous and discrete wavelet transforms on the speed signal, and the kurtosis index is used to detect the fault. The advantages of this method are the reduction of the cost of fault detection due to the non-use of additional sensors, the high robustness of the detection method to noise and disturbances due to proper insulation between the Hall effect sensors and the power circuit, and the capability using of this method for the stationary and non-stationary states. In [50], a new approach for detecting demagnetization in a linear PM synchronous motor is investigated. This method measures the magnetic field density at three points using three Gaussian probes. Then, the severity, location, and type of faults are detected using the machine learning-based detection algorithm. This method has a valuable ability for non-operating situations. Also, the machine learning method is compared with the Backpropagation Neural Network (BPNN) and Probabilistic Neural Network (PNN) in this study, and its superiority is determined.

To conclude, it is found that the main problem of methods to measure the magnetic field directly is not-online detecting and hurting the structure of PMSM because of the disassembling of the machine. However, analytical methods are proposed to calculate and estimate the magnetic field using other machine signals. These methods have different benefits and drawbacks, and some do not apply to different machines because of structure. For instance, the BLDC machines have a Hall effect sensor in their structure; therefore, the magnetic field is monitored without any disassembling.

2.1.2.5 Demagnetization Detection Using the Signal Injection

The signal injection-based methods for detecting demagnetization are used in PM machines. As shown Figure in 2-4, detecting faults of the stator winding, in which by injecting a high-frequency signal, the rate of increase of the machine inductance is investigated. Parameter estimation methods using signal injection solve the problem of model sensitivity to temperature and saturation and estimate the severity of the fault. This method can be applied to all machines powered by an inverter. Machine parameters are estimated in the signal extracting process, and an essential indicator for fault detection is extracted [51].

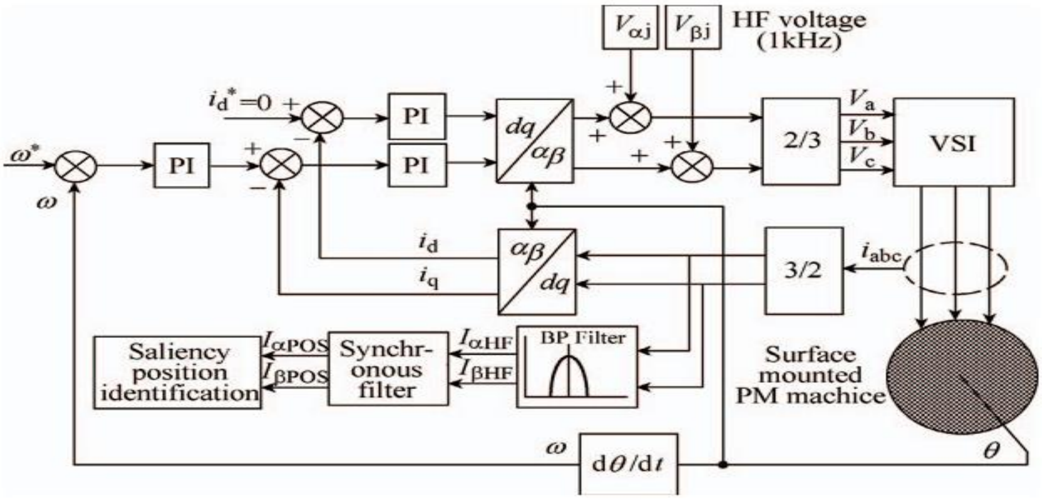


Figure 2-4: Applying a high-frequency signal for detection [52]

By injecting current into the faulty machine, the effect of demagnetization on the PMs is reflected in a meaningful framework in the machine signals. Figure 2-5 shows the changes $L_d{}'$ compared to I_d for the healthy state, demagnetization, and eccentricity [53]. High sensitivity, non-dependence on uncertainty and operating conditions, no need for installing another measurement equipment, and low cost are some of the prominent features of this approach. Although, continuous monitoring of the PM situations is not possible. Therefore, this method is only suitable for machines that are periodically inspected.

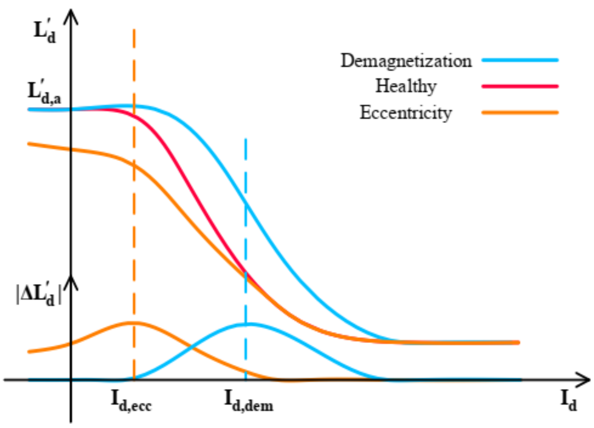


Figure 2-5: Inductance derivative changes in the d-axis direction versus to the current in the d-axis direction [53]

2.1.2.6 Demagnetization Detection Using the Torque Signal

The electromagnetic torque of the PM motor is caused by the interaction of the rotor magnetic field and the stator winding field in the air gap. Demagnetization changes the flux density distribution of the air gap, which in turn creates torque ripples and increases the frequency components. The listed frequency components are more strongly dependent on the severity of the demagnetization and the rate of change of these components.

Similar to frequency analysis methods, these methods have similar problems. For this purpose, a time characteristic for detecting the torque signal is proposed in [54]. One of the methods that can detect hidden patterns in torque time data is the Time Delay Embedding (TDE) method. The torque profile is transferred to a reconstructed profile called fuzzy space. Increasing the severity of demagnetization reduces the rate of change of torque. In addition, the estimated radius of rotation points around the center of mass can be a good criterion for fault detection. This radius is:

$$\mathrm{R_g} = \sqrt{\frac{\sum_{k=l+1}^{\mathrm{M}} \mathrm{d}(k)^2}{M-l}} \tag{2-17}$$

In this regard, l is the time delay, d(k) is the distance between the center and the point k and M is the number of observations. Figure 2-6 shows that the rotation radius decreases with increasing demagnetization. Direct torque measurement requires expensive sensors and is also an aggressive method. Therefore, this method attempts to estimate electromagnetic torque in the presence of saturation using accurate analytical equations.

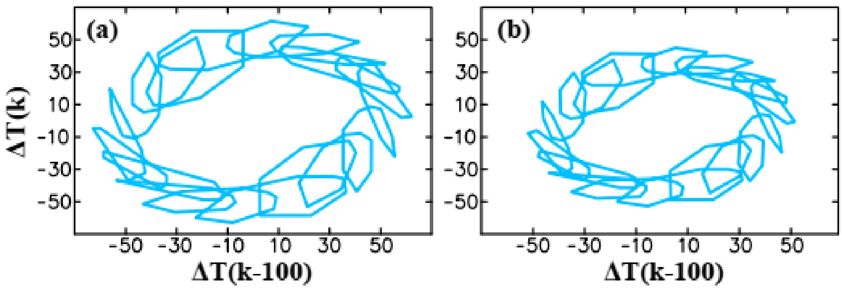


Figure 2-6: Fuzzy space (a) Healthy state (b) 25% demagnetization state [54]

In [55], a torque ripple-based method for monitoring the state of the linkage flux is introduced to detect demagnetization in the stationary and non-stationary states. It is proved that demagnetization has a significant effect on the output torque. The torque ripples are investigated by examining the torque characteristics and extracting the energy order using a Kalman filter. In this research, it is found from the experimental results that the torque ripple in the order of three energies shows the changes of the torque ripple in the best way. Finally, the Bayesian algorithm uses these characteristics to estimate the linkage flux.

In [56], a method based on torque angle measurement is proposed to detect turn-to-turn and demagnetization faults. One of the most important features of this method is the ease of implementation. The proposed method applies to Field Oriented Control (FOC) converters and does not require using additional sensors. In addition, because of its high sensitivity to other faults and disturbances, this method does not require the use of large storage memory for implementation. Besides, an online demagnetization detection method for using torque ripple is proposed in [57]. First, by using continuous wavelet transform (CWT), the system noise is reduced, and then by applying edge energy spectrum analysis, the dynamic energy of the torque ripple is revealed. Finally, the demagnetization is estimated using the Grey System Theory (GST) algorithm.

To conclude, based on reviews of different fault detection methods in this chapter, it is found that the model-based methods are proper because of their ability to analyze the machine using the extracted model with low time-consumption. However, the main problem for accurate analysis is the non-availability of all machine models to present a comprehensive method. The data-driven methods are such popular methods, especially in the industrial application, because of the availability of analysis data online. However, it is presented that each technique for each signal has different positive and negative points. Compared to other methods, deep learning methods have an exceptional ability to analyse large-scale data.

The lack of labelled data is one of the most important challenges in working on fault detection algorithms. According to the review of different deep learning methods for detecting faults, it is found that methods such as VAE and GANs are generative methods that it is possible to generate new data in their process. Besides, VAE is known as one of the most valuable methods for extracting features [38]. Therefore, preparing one algorithm based on VAE for various types of sensors can be a valuable idea for monitoring electrical machines. Also, one of the strengths of VAEs is their robustness to noisy or missing data. Unlike traditional autoencoders, VAEs are designed to generate new data samples by sampling from the learned compressed representation. This sampling process allows VAEs to generate realistic samples even when the input data is incomplete or contains noise. Moreover, VAEs can be trained with a wide range of loss functions, making them adaptable to a variety of data types and applications. Overall, the robustness of VAEs makes them a versatile and useful tool for many machines learning tasks.

2.2 Thesis Objectives

This thesis aims to provide a new way to diagnose demagnetization fault in the PM synchronous machines. On the one hand, this method should analyze a considerable amount of data with low process time. On the other hand, it should apply to different signals in different operating conditions. Besides, this method should have a proper ability for detecting various faults with high accuracy. Also, data fusion is another crucial challenge to investigate in this thesis using the new method. Therefore, it is decided to extend research on VAE for different scenarios and signals. The most important goals of this research work are prepared as follows:

- Extracting the current, vibration, and sound signals for different operating conditions;
- Implementing the proposed method on various faults consisting of demagnetization, bearing, and eccentricity faults;
- Investigating the performance of the proposed method for sensor fusion;
- Evaluating the ability of this method for combinations of faults;
- Comparing the proposed method to the several popular machine learning methods in different aspects such as accuracy and training time;
- Investigating the influence of pre-processing methods on increasing the ability of the proposed method

3 CONCEPTUAL AND THEORETICAL FOUNDATION

The main goal of this chapter is to explain the theory of the proposed method for the diagnosis of faults in the PMSMs. In the first section of this chapter, machine learning is introduced, and its different classes of methods comprising supervised, unsupervised, and reinforcement learning are explained thoroughly. Then the importance of using deep learning methods is presented. Besides, some of the important machine learning methods are explained. In the next step, the autoencoder is investigated, and the theory of VAE is scrutinized entirely.

3.1 Machine Learning

Machine learning is a subset of artificial intelligence based on learning and doing tasks according to achieved experiences during the training process. In other words, the primary goal of using machine learning methods is to give machines tasks, and do similar or new tasks according to achieved experiences. This process is similar to learning by humans. Therefore, machine learning has a significant impact on improving productivity and automation of tasks. However, researchers were not initially not able to program the machine to perform more complex tasks, except for a few simple tasks such as finding the shortest path between points A and B [58].

Arthur Samuel introduced the concept of machine learning in 1959, which he defined as "the field of study that gives computers the ability to learn without being explicitly programmed" [59]. According to the importance of this method to enhance the efficiency, and ease of doing work, it has been developed. Nowadays, these techniques are used in many technology fields, and its use has become so widespread that people are often unaware of its presence in nowadays tools. At first, it was thought that it is only possible to find complex patterns in data by the human brain, but with the advent of machine learning methods in which the analysis is similar to the human brain, this view has changed.

Machine learning algorithms utilize a set of data referred to as the training data to develop and construct the necessary models. These models are capable of making predictions on the behaviour of a given system based on the data used to create them. Upon the introduction of new data referred to as the testing data to the algorithm, the resulting predictions are meticulously evaluated alongside their corresponding labels.

If the evaluation metrics confirm that the model performs satisfactorily on the testing data, the machine learning algorithm is deemed to be successful. Figure 3-1 shows that the machine learning method is divided into supervised, unsupervised, and reinforcement learning. Machine learning empowers data scientists to analyze large amounts of data, automate processes, and make awareness in the least amount of time.

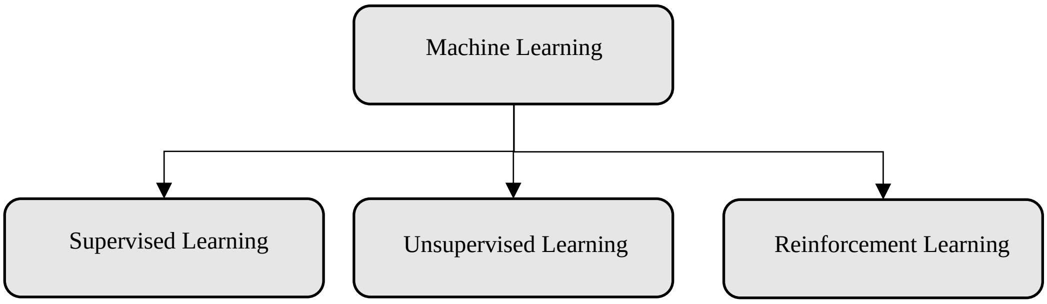


Figure 3-1: Machine learning sub-branches

3.1.1 Supervised Learning

The supervised learning method is similar to teaching students under supervision and guidance. A data set is prepared that acts as a teacher and is responsible for teaching the machine or model. Once the relevant model is trained, it makes the necessary predictions and decisions about new data entering the system [60]. As shown in Figure 3-2, the data in the supervised learning method is divided into three sections; train data, test data and validation data; usually 70 percent of data is dedicated to training data and 30 percent of them is dedicated to testing and validating. During the machine learning process, the training data is utilized to tune the hyperparameters of each method. The testing data is then used to assess the performance of these methods and determine if any additional parameter adjustments are necessary. Once the performance of the selected method is deemed satisfactory on the testing data, a separate dataset known as validation data is used to validate the method's performance. Some of the significant methods of supervised learning can be listed as follows:

- K-Nearest Neighbourhood
- Decision Tree
- Random Forest
- Neural Networks

- Logistic Regression
- Support Vector Machines
- Naïve Bayes

In this thesis, k-nearest neighbourhood, decision tree, and naïve Bayes are selected for evaluation using collected data, and the results are presented in Chapter 5 for comparing the proposed method and popular machine learning methods. In this chapter, they are applied to the crucial supervised learning technique presented above.

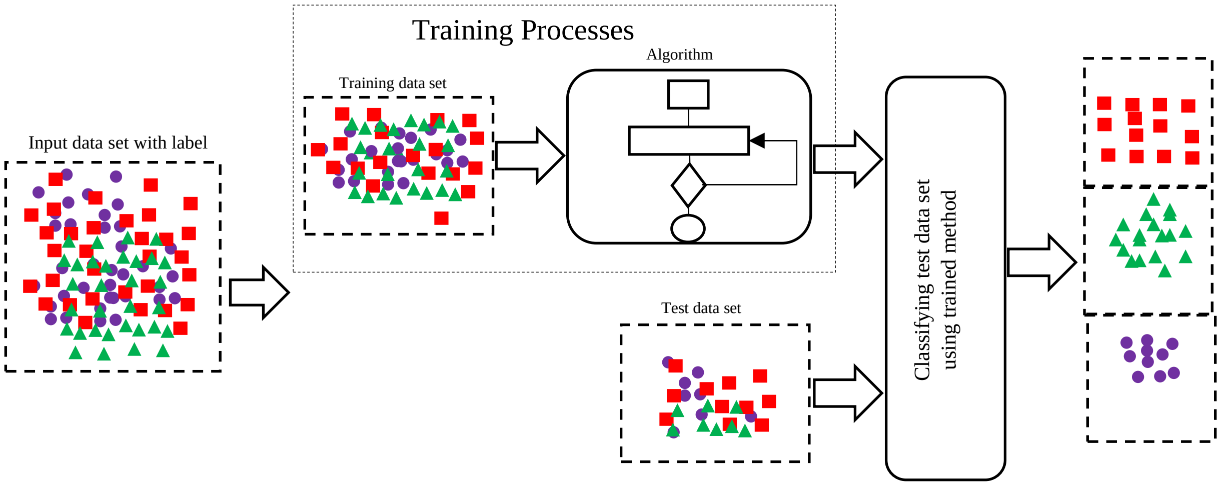


Figure 3-2: The supervised learning procedure

3.1.1.1 K-Nearest Neighbourhood

K-Nearest Neighbourhood (KNN) is one of the simplest and applicable supervised machine learning algorithms in regression and classification problems. This method determines the value of each new data by the average of its k neighbours' values in regression. Also, the class of each new data in classification is determined using the majority classes of its k neighbours. According to importance of classification for this thesis, the explanation of KNN is based on classification in [61]. The KNN method is effective, nonparametric, and easy to implement. k is a small, odd, and positive value, and the class of each element of a given sample is determined from the classes of the k nearest neighbours, where being "near" can be validated via a problem-depending metric (see the example below). However, the classification time consumption is long, and it is difficult to find the optimum value of k. The best choice of k is determined depending on the distribution of data. In general, increasing the value of k reduces the effect of noise on the classification; however, it blurs the boundaries between the classes. Figure 3-3 shows an example of a KNN classification algorithm using multidimensional property vectors in which the data classes are illustrated by different shapes such as triangles, circles, and squares. Also, the new sample is determined using a diamond. For instance, if k=2, the diamond class is determined based on the classes of the two nearest samples. Therefore, the distance between the diamond (test sample) and all of the samples is calculated, and based on the two closest data in one class, the class of this sample is defined. According to distance calculation, the test sample belongs to the circle class. Therefore, the training step for the KNN is based on the values and classes of the k-nearest neighbours. Euclidean distance and Cosine similarity are usually used as a measuring tool in this method.

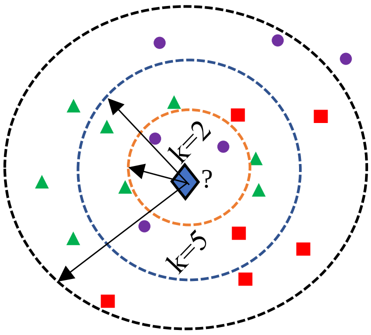


Figure 3-3: Classifying data using KNN

In machine learning, a common approach to classification tasks involves representing training samples as vectors in a possibly high dimensional feature space. Each element of a sample is associated with a particular class, and during testing, the elements of a new sample are also represented as vectors in the feature space. To classify an element of the new sample, the nearest k elements from the training sample to the test specimen are identified based on their Euclidean distance, cosine similarity, or other appropriate metric. The class label of the new specimen is then assigned based on the majority class label of the k nearest elements of the training sample. It is worth noting that there are various approaches to classification, and the k-nearest neighbor algorithm is just one of them. Nonetheless, this method is often used due to its simplicity and effectiveness, especially when dealing with small to medium-sized datasets. The three critical factors in the k-nearest neighbor algorithm are:

- Distance or Similarity Criterion: This factor determines the similarity or dissimilarity between the feature vectors of the training and test samples based on a chosen metric like Euclidean distance or cosine similarity.
- Number of Close Neighbors: This factor represents the number of closest training samples that are used to determine the class of the test sample, and it can have a significant impact on the accuracy of the classification.
- Rule for Determining Class: This factor determines the class label of the test sample based on the majority class label of the k nearest training samples, and it is a crucial component of the algorithm.

3.1.1.2 Decision Tree

One of the most popular machine learning methods is the Decision Tree (DT). Similarly to KNN, it is counted as a predictive method for classification data in machine learning [62]. A classification tree is used to classify a dataset commonly in marketing, engineering, and medical activities. In a decision tree structure, tree-derived predictions are described by rules. As shown in Figure 3-4, each path from the root to the leaf of the tree expresses a rule. Then the leaf is labelled with the class to which it belongs. The main components of the decision tree are defined as follows:

- Leaf Nodes: Nodes where successive divisions end. The leaf is dedicated to a label;
- Root: The root node of the tree;
- Branches: In each internal node, as many branches as possible are created;

To construct a decision tree, one typically begins with a dataset and selects a feature that best divides the data into groups. The Gini index or information gain is then calculated for each possible split to determine which one produces the highest quality tree. The Gini index measures the impurity of a dataset, where a value of 0 means that the data is perfectly homogeneous and all elements belong to the same class, and a value of 1 means the data is perfectly impure and equally distributed among all classes. It is a proper method for categorical variables or binary classification. The Gini index is calculated as follows:

$$\mathrm{Gini} = 1 - \sum_{i=1}^{j} \mathrm{P}(i)^2 \tag{3-1}$$

where, $\mathrm{P}(i)$ represents the proportion or probability of an event i occurring in a set of possible events, and j represents the number of groups or categories being considered. Besides, information gain is an index which is determined based on the entropy values. Entropy shows the lack of information in one node. Entropy and information gain are calculated as follows:

$$\mathrm{Entropy} = -\sum_{i=1}^{n} \mathrm{P}_i \log_2(\mathrm{P}_i) \tag{3-2}$$

$$\text{Information Gain} = \text{Entropy Before splitting} - \text{Entropy after splitting} \tag{3-3}$$

where P_i represents the proportion or probability of an event i occurring in a set of possible events, and n represents the number of possible outcomes or categories. The information gain increases, when the entropy is decreased after splitting on a sub-branch. The best node for splitting is a node with high information gain. The decision tree algorithm starts by selecting the entire dataset as the root node and identifying the feature with the highest Gini index or information gain as the splitting criterion for the current node. The data is then split into two or more branches based on the selected feature's values, and the process is repeated recursively for each branch until a stopping criterion is met, such as a minimum number of samples in a leaf node or a maximum depth of the tree. Once the tree is built, new data points can be classified by traversing the tree from the root node to a leaf node and assigning the target variable based on the majority class in that node.

Pruning the decision tree is on the opposite direction of dividing. In other words, pruning is a removing process with sub-nodes in the decision tree. When a decision tree is created, a number of branches reflect anomalies in the training data that result from outliers or noisy data. In some tree algorithms, pruning is a part of the algorithm. While in others, pruning is used only to solve the problem of overfitting. There are several methods using statistical criteria to remove less reliable branches. Pruned trees are smaller and less complex, making them easier to understand. They are usually faster and better at not sorted test data than unpruned trees.

Generally, there are two approaches to tree pruning: pre-pruning and post-pruning. In the pre-pruning approach, a tree is pruned by repeated stops in the early stages of tree construction. As soon as a stop is made, the node turns into a leaf. The post-pruning approach is a more typical decision tree in that it removes sub-trees from a fully-grown tree. A subtree in a node is pruned by removing the branches and replacing them with a leaf.

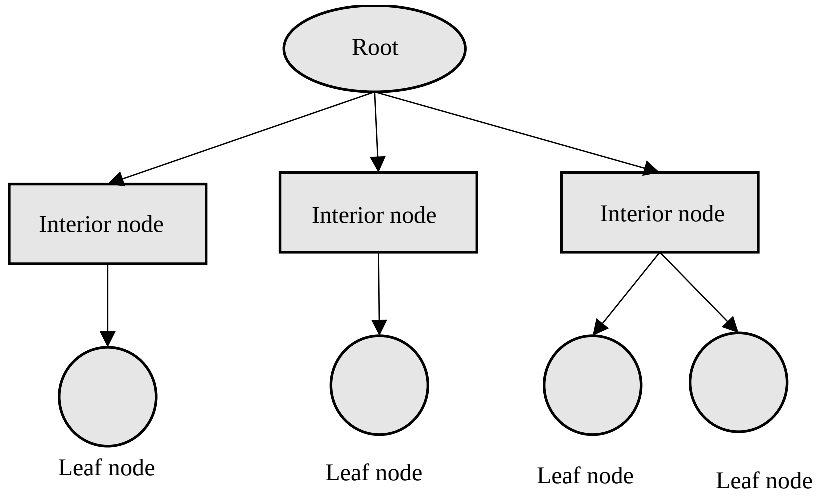


Figure 3-4: Decision tree structure

3.1.1.3 Naïve Bayes

Naïve Bayes (NB) classifier based on the Bayes theorem, provides a way to calculate the probability of a hypothesis based on its prior probability [63]. It's called "naive" because it makes a simplifying assumption that the features of a data point are independent of each other, given the class label. In general, Bayes' theorem involves observed data D and a hypothesis h, belonging to the H hypothesis space, that needs to be investigated. The Bayes theorem reads as follows [63]:

$$\mathrm{P}(h|D) = \frac{\mathrm{P}(D|h)\mathrm{P}(h)}{\mathrm{P}(D)} \tag{3-4}$$

where $\mathrm{P}(h)$, $\mathrm{P}(D)$, $\mathrm{P}(h|D)$, $\mathrm{P}(D|h)$ are the prior probability that hypothesis h is true, the prior probability of observing data point D, the posterior probability that the hypothesis h is true given the data D, and the likelihood function that specifies the probability of observing data point D given that hypothesis h is true, respectively. The maximum a posteriori (MAP) hypothesis is defined to be the maximally probable hypothesis as follows:

$$h_{\mathrm{MAP}} = \underset{h \in H}{\operatorname{argmax}} \mathrm{P}(h|D) = \underset{h \in H}{\operatorname{argmax}} \frac{\mathrm{P}(D|h)\mathrm{P}(h)}{\mathrm{P}(D)} \tag{3-5}$$

where P(D) is removed because of its constant quantity, and independence from h. Finally, the equation (3-6) is extracted as follows:

$$h_{\mathrm{MAP}} = \underset{h \in H}{\operatorname{argmax}} \mathrm{P}(D|h)\mathrm{P}(h) \tag{3-6}$$

where h_{MAP} is the most probable hypothesis in the H space, which is probable to happen. In many domains, the prediction accuracy of the NB classifier compares well with other complex learning algorithms, including decision tree learning, rule learning, and instance-based learning algorithms. The NB classifier is based on the target function $f{:}x \rightarrow h_j$. This function maps data point x described by the tuple of features values $(a_1, a_2, \dots, a_{i,} \dots, a_n)$ to the hypothesis h_j. The MAP hypothesis for each feature (a_i) is calculated as follows:

$$h_{\mathrm{MAP}} = \underset{h_j \in H}{\operatorname{argmax}} \mathrm{P}(a_i|h_j)\mathrm{P}(h_j) \tag{3-7}$$

where $\mathrm{P}(h_j)$ and $\mathrm{P}(a_i|h_j)$ are estimated based on their frequency over the training data. As the features are conditionally independent, the h_{NB}, MAP hypothesis for each data point (x), is presented as follows:

$$h_{\mathrm{NB}} = \underset{h_j \in H}{\operatorname{argmax}} \mathrm{P}(h_j) \prod_{i=1}^{n} \mathrm{P}(a_i|h_j) \tag{3-8}$$

3.1.2 Unsupervised Learning

Unsupervised learning is dedicated to a subset of machine learning methods which there are no labelled data. Therefore, the supervised learning method is not applicable for this type of data, which often occurs for industrial datasets. The important features are extracted through trained observations, as shown in Figure 3-5. When a dataset is entered to the algorithm, the model automatically recognises the patterns in the data using clustering or other methods. Unsupervised learning is more challenging than supervised learning because of the difficulty of separating classes without labels. K-means, hierarchical clustering methods, and Self-Organising Maps (SOM) are some of the popular unsupervised learning methods [63].

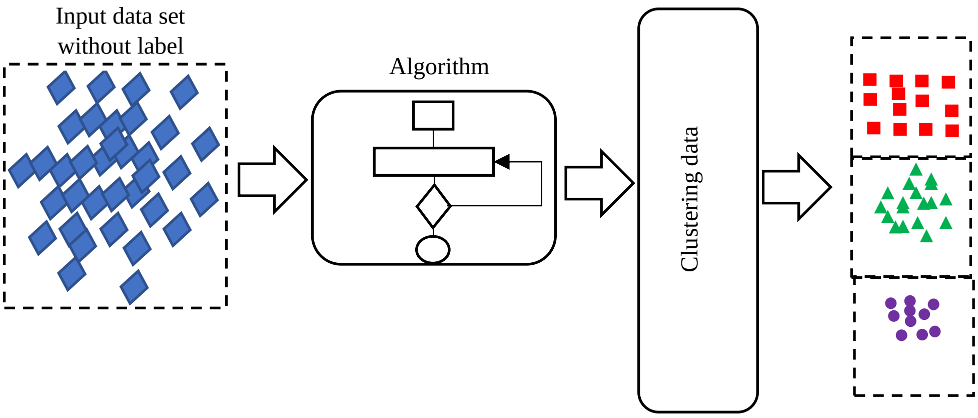


Figure 3-5: The unsupervised learning

3.1.3 Semi-supervised Learning

This category lies between the supervised and unsupervised machine learning methods. In case of semi-supervised learning, the larger part of the provided data are unlabeled data. Therefore, the process of training is done using labelled and unlabeled data. Besides, the labelled data is used to train the network similar to supervised algorithms, and unlabeled data is used to train algorithms using unsupervised learning methods. As a consequence, the process of training is more complicated than for the aforementioned methods. There exist a lot of industrial case studies in which this situation occurs and the research and development face a lack of labelled data for the diagnosis of faults.

3.1.4 Reinforcement Learning

Reinforcement learning, one of the most applicable paradigm of machine learning in robotics, is based on learning by testing different states [58]. In other words, an agent is taught to operate according to its achieved experiences. This concept is also known as the trial and error method. This agent is rewarded or penalized based on the right or wrong operation. Finally, the model achieves the ability to improve through positive and negative scores based on its operation in different states. The training process is parallel with testing until the new input data is entered. Q-learning and Temporal Difference are some of the popular reinforcement learning methods.

3.1.5 Feature Extraction Methods

Feature extraction is one of the most important stages of classification and regression using machine learning methods [64]. The feature extraction process contains extracting and selecting features. If labelled data is available, the feature extraction method is based on supervised learning for choosing the most discriminant and valuable features. However, unsupervised learning methods are used for extracting features for unlabelled data. In the unsupervised and semi-supervised feature extraction methods based on the existence of unlabelled data, the correlation matrix is used to investigate the similarity of the features. In the continuation of this research work, two applicable feature selection methods are used: Principal Component Analysis (PCA) as a supervised feature selection and Linear Discriminant Analysis (LDA) as an unsupervised feature extraction method.

3.1.5.1 Principal Component Analysis

In [65], PCA is proposed as a method to produce new variables using linear combinations of uncorrelated variables by Karl Pearson in 1901. This method is known as a feature extraction method which reduces dimensionality parallelly. Generally, feature extraction is a mapping from an input data space to another space, in which the separability is better in the first one. Therefore [63]:

$$G \epsilon R^{p\times d}: x \epsilon R^{p} \rightarrow y = G^{T} x \epsilon\, R^{d} (d \ll p) \tag{3-9}$$

where, G is a matrix that transfers the data from the space with p features to another space with d features. Therefore, a reducing dimensionality method maps from a space with $p\times n$ dimension to a space with a $d\times n$ dimension. The primary objective of PCA is to identify the eigenvectors and eigenvalues of the covariance matrix of a dataset, in order to determine the optimal vectors along which data can be projected to maximize separability between classes. The data set can be defined as follow:

$$\mathrm{X} = \{x_{11}, \dots, x_{pn}\} \in R^{p\times n} \tag{3-10}$$

In equation (3-10), p is the number of features, and n is the number of row data. The first principal component (Z_1) is defined by applying a linear transformation as follows:

$$Z_1 = {a_1}^{T}\mathrm{X} = \sum_{i=1}^{p} a_{i1} x_i \tag{3-11}$$

where x_i represents the value of i^{th}feature for a specific observation in the dataset X, and a_{i1} represents the coefficient assigned to feature x_i.The variance of first principal component (Z_1) is defined as follows:

$$\begin{aligned}\mathrm{var}[Z_1] = \mathrm{E}((Z_1 - \overline{Z_1})^2) &= \frac{1}{n}\sum_{i=1}^{n}(a_1{}^{\mathrm{T}}x_i - a_1{}^{\mathrm{T}}\bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} a_1{}^{\mathrm{T}}(x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}a_1\end{aligned} \tag{3-12}$$

If $\mathrm{S} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$, the formation of equation (3-12) changes to (3-13) as follows:

$$\mathrm{var}[Z_1] = a_1{}^{\mathrm{T}}.S.a_1 \tag{3-13}$$

Also, each data set can be defined as $X = U\Sigma V^T$. Here, U, V, and Σ are the matrices of eigenvalues, eigenvectors and covariance, respectively. For finding a_1 that maximize $\mathrm{var}[Z_1]$, the Lagrange method can be a proper way. With the restriction $a_1{}^{\mathrm{T}}a_1 = 1$, the Lagrange function is defined as follows:

$$L = a_1{}^{\mathrm{T}}Sa_1 - \lambda(a_1{}^{\mathrm{T}}a_1 - 1) \tag{3-14}$$

For finding the critical points of L, a derivative should be implemented on the Lagrange equation in equation (3-14), Therefore:

$$\frac{\partial L}{\partial a_1} = Sa_1 - \lambda a_1 = 0 \Rightarrow (S - \lambda I_1)a_1 = 0 \tag{3-15}$$

$$\lambda_1 = a_1{}^{\mathrm{T}}.S.a_1 \tag{3-16}$$

So, the first vector (a_1) is found according to equation (3-16). For finding the next one, the Lagrange equation should be extended, and calculated again, it is important to note that $\mathrm{var}[Z_1]$ and $\mathrm{var}[Z_2]$ should be completely uncorrelated. Thus, the Lagrangian can be defined for finding the next principal component (a_2) as below:

$$\begin{aligned}&L = a_2{}^{\mathrm{T}}Sa_2 - \lambda(a_2{}^{\mathrm{T}}a_2 - 1) - \varphi a_2{}^{\mathrm{T}}a_1 = \frac{\partial L}{\partial a_2} = Sa_2 - \lambda a_2 - \varphi a_1 = 0 \\ &\Rightarrow (S - \lambda I_1)a_2 - \varphi a_1 = 0\end{aligned} \tag{3-17}$$

In the final result of equation (3-17), φ should be zero, and the next eigenvalues can be found as follows:

$$\lambda_2 = a_2{}^{\mathrm{T}}.S.a_2 \tag{3-18}$$

Therefore, it is clear that for finding the principal component of data, it is sufficient to find the eigenvectors according to eigenvalues from biggest to lowest. Also, all the components are entirely uncorrelated. Generally, it is possible to define it as follows:

$$\mathrm{var}[Z_k] = {a_k}^T.S.a_k = \lambda_k \tag{3-19}$$

Therefore, the k^{th} principal component is the k^{th} eigenvector of covariance matrices of data. As shown in Figure 3-6, based on the distribution of data, the first eigenvector is found then the next vector is placed perpendicular to the first one, which shows the uncorrelation between these vectors. As can be seen in Figure 3-6, the separation of data by implementing PCA is clearly better than the first space.

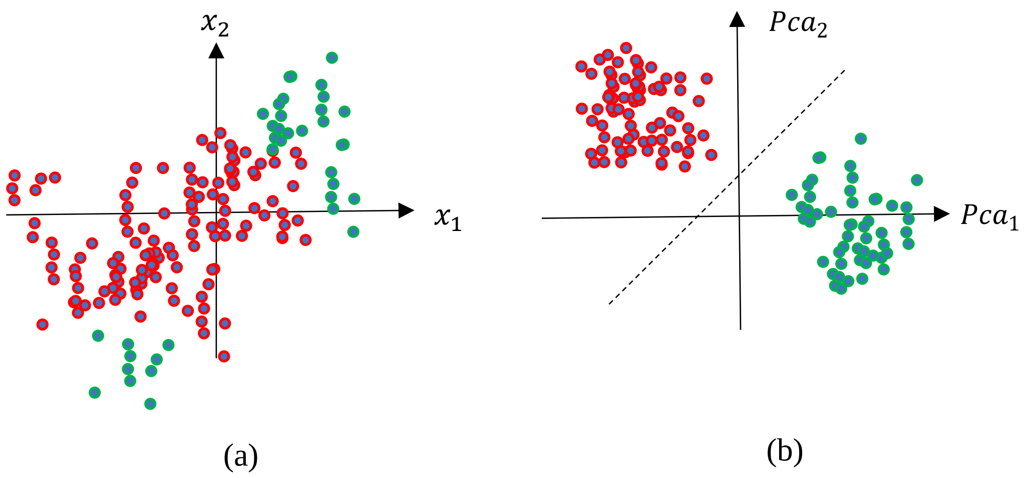


Figure 3-6: Implementing of PCA on data (a) Plotting data using two features (b) Plotting data after implementing PCA using the first two components

3.1.5.2 Linear Discriminant Analysis

The theory of Linear Discriminant Analysis (LDA) or Fisher Discriminant Analysis (FDA) was proposed by Ronald Fisher in 1936 [66]. The main goal of LDA is finding the best line for projecting data on this line with higher separability compared to the first space. LDA is accounted for as a supervised learning method, unlike PCA is accounted for as an unsupervised learning method. As shown in Figure 3-7, the separability on line L1 is better than on line L2.

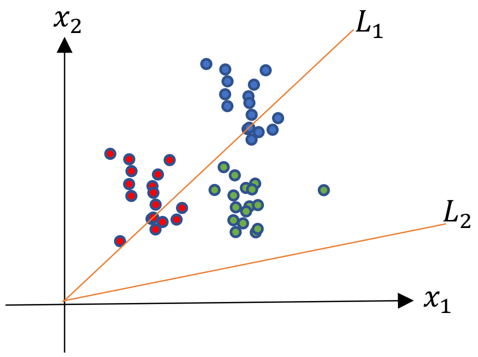


Figure 3-7: Two lines for projecting data on them

The first step in LDA is defining one objective function in which the distance between the mean of classes is high and the variances of data are low. Therefore, if the projection of data on the intended line is defined as y, the equation for this transformation is presented as follows:

$$y = W^T x \tag{3-20}$$

where W is a transformation matrix to project the data on the new line. Also, the essential parameters of the distribution of each class data are defined as follows:

$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \ , \ \tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{y\in\omega_i} W^T x = W^T \mu_i \tag{3-21}$$

where μ_i represents the mean of data belonging to the class ω_i, and $\tilde{\mu}_i$ is the mean of data on the projected line belonging to the class ω_i. Also, N_i is the number of data that existed in the class ω_i. Also, the scatter matrix $(\tilde{s}_i{}^2)$ for each class can be defined as follows:

$$\tilde{s}_i{}^2 = \sum_{y\in\omega_i}(y-\tilde{\mu}_i)^2 = \sum_{y\in\omega_i}(W^T x - W^T\mu_i)^2 = W^T \sum_{y\in\omega_i}(x - \mu_i)(x-\mu_i)^T W = W^T S_i W \tag{3-22}$$

where S_i is :

$$S_i = \sum_{y\in\omega_i}(x-\mu_i)(x-\mu_i)^T \tag{3-23}$$

The objective function of LDA can be defined for any number of classes. However, to simplify the explanation, the objective function for two classes is defined as follows:

$$j(W) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1{}^2 + \tilde{s}_2{}^2} \tag{3-24}$$

According to the numerator and denominator of equation (3-24), increasing the distance between mean's data and decreasing the variances data causes to maximize the objective function. The numerator of the equation (3-24) can be presented as:

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (W^T\mu_1 - W^T\mu_1)^2 = W^T(\mu_1 - \mu_2)^2 W = W^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T W = W^T S_B W \quad (3\text{-}25)$$

where S_B is the between class scatter matrix. The denominator of equation (3-24) is defined as:

$$\tilde{s}_1{}^2 + \tilde{s}_2{}^2 = W^T S_W W \quad (3\text{-}26)$$

where S_w is the within class scatter matrix. Therefore, the formation of equation (3-24) can be changed to:

$$j(w) = \frac{W^T S_B W}{W^T S_W W} \quad (3\text{-}27)$$

For finding the transformation matrix W for maximizing the objective function in equation (3-27), it is needed to solve the following equation [66]:

$$\frac{\partial}{\partial W}[j(W)] = 0 \Rightarrow \frac{\partial}{\partial W}\left[\frac{W^T S_B W}{W^T S_W W}\right] = W^T S_W W \frac{\partial}{\partial W}(W^T S_B W) - W^T S_B W \frac{\partial}{\partial W}(W^T S_W W) = W^T S_W W(2S_B W) - W^T S_B W(2S_W W) = \frac{W^T S_W W(2S_B W)}{W^T S_W W} - \frac{W^T S_B W(2S_W W)}{W^T S_W W} = S_B W - jS_W W = (S_W{}^{-1} S_B - jI)W = 0 \quad (3\text{-}28)$$

Therefore, the transformation matrix W for reaching the proper separation is the eigenvectors of $S_W{}^{-1}S_B$.

3.1.6 Artificial Neural Network

Artificial Neural Networks (ANNs) are computational systems that are originally inspired by biological neural networks in human beings [67]. Biological neural networks make up the nervous system that makes up the brains of many animal species. Artificial neural networks were first introduced in the 1940s, and given the ability of this algorithm, it becomes known as one of the most widely used machine learning methods [68]. These networks have been greatly expanded in recent years, and deep neural networks have been introduced [69]. Deep neural networks are much more complex than neural networks, and research has shown that these methods have an extraordinary ability to solve complex problems such as classification problems. This section is prepared based on [70, 71].

The main goal of ANN is to learn one process according to input and output data. Each ANN has different components: neurons, an input layer, an output layer, and one or more hidden layers. The learning process needs one loss function to determine the goal of training, an optimizer such as Stochastic Gradient Descent (SGD) for finding the optimised value for each coefficient consisting of weights and bias, and one method for updating coefficients of the network such as backpropagation. It can be noted that when the number of hidden layers increases from one layer to more, the neural network is called deep neural network. A series of transformation can be presented for describing the process of an ANN as shown in Figure 3-8. The first transformation involves performing linear combinations of M from the input vector, described as $x = (x_1, x_2, \dots, x_D)$ and M is the size of hidden layer. Therefore, the value assigned to neuron z_j in the hidden layer for iteration is presented as follows:

$$z_j{}^{(n)} = f\left(a_j{}^{(n)}\right) = f\left(\sum_{i=1}^{D} W_{j,i}{}^{(n)} x_i + W_{j,0}{}^{(n)}\right), \quad \forall j = 1, \dots, M \tag{3-29}$$

where, $W_{j,i}{}^{(n)}$ is the weight between input data (x_i) and neuron (z_j) in iteration number n, $W_{j,0}{}^{(n)}$ is the bias coefficient for z_j in iteration number n, and f is an activation function [70]. Then this process is continued for the next hidden layer in a deep neural network with several hidden layers. Finally, the data are transferred to the last layer which is called output layer. The k^{th} output ($y_k{}^{(n)}$) for iteration number n is defined as follows:

$$y_k{}^{(n)} = h\left(\sum_{j=1}^{M} W_{k,j}{}^{(n)} z_j{}^{(n)} + W_{k,0}{}^{(n)}\right), \quad \forall k = 1, \dots, K \tag{3-30}$$

where, $W_{k,j}{}^{(n)}$ is the weight between neuron (z_j) and output (y_k) in iteration number n, $W_{k,0}{}^{(n)}$ is the bias coefficient for y_k in iteration number n, and h is an activation function.

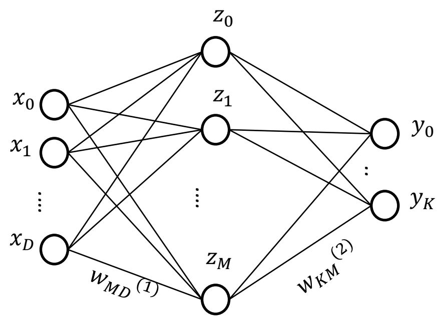
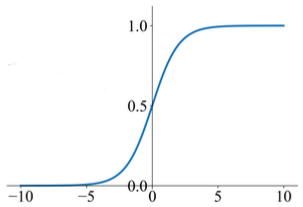


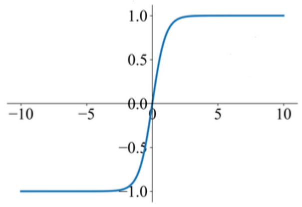
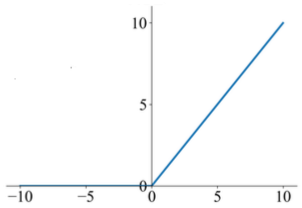
Figure 3-8: Neural network’s structure

3.1.6.1.1 Activation Function

Activation functions are actually like the gate that exists in each neuron. The input of this gate is the input of each neuron in each layer and its output is transferred to the next layer. The activation function decides how each neuron is activated or not. In a simple neural network, using activation functions is not essential; however, in neural network with high number of layers using activation functions can be helpful to get better results. Besides, when non-linearity is added to a network, activation functions are needed. There are different activation functions for different applications. In this section, some of the most applicable are presented as shown in Table 3-1. The ReLU function is so applicable in hidden layers which return zero if the value of input data is negative and return input data otherwise. It is found that its performance in process of training neural network is better than sigmoid and hyperbolic tangent activation [70]. The SoftMax function is one of the most applicable activation functions in the output layer for multiclass classification which restricts the outputs of each unit in the range [0, 1]. Thus, the sum of the total outputs is equal to 1.

Table 3-1: Activation functions equations and graphs

Activation function	Equation	1D-Graph
Sigmoid activation function	$\sigma(x) = \dfrac{1}{1 + e^{-x}}$	

Hyperbolic tangent activation function	$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$	
Rectified Linear Unit (ReLU) activation function	$\mathrm{ReLU}(x) = \max(x, 0)$	
SoftMax activation function	$\mathrm{SoftMax}(x_j) = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}},\ j = 1, \dots, K$	-

3.1.6.1.2 Training Algorithm

One of the fundamental parts of each neural network is the training algorithm for updating all of the weights and biases to get a proper result. In [72], a learning algorithm is proposed for the binary classification based on the perception convergence theorem. In this method, the coefficients are updated by finding the upper and lower bounds. In [73], the ADALINE method is proposed and it is extended based on the Least-Mean-Square (LMS). These methods were the first method that proposed for single-layer perceptron neural networks; however, they are not applicable for multilayer perceptron neural networks. Therefore, the backpropagation method based on investigating a propagating error in the opposite direction in a feedforward process was proposed in the 1970s.

In 1974, the backpropagation technique was proposed for solving the problem of how to train the multilayer network for the first time. This method was examined and it was found that this method is much faster than other methods. Then in 1980s, one framework for this method was proposed in [74]. This framework later was used as an initial point to start the theory of deep neural networks. The basis of this method is following an iterative optimization technique, in which in each iteration based on an objective function the weights and bias coefficients are updated. It can be noted that objective function can be called loss or cost function. The loss function can be defined as follows:

$$J = y - y' \tag{3-31}$$

where $\boldsymbol{y'}$ is the estimated, and $\boldsymbol{y}$ is the true value of the target. A lower value of the loss function indicates a better estimation. Using this loss function, the optimized amount for different parameters of the network can be extracted as:

$$J = \min(\mathrm{f}(x, \theta)) \tag{3-32}$$

where θ is the weight coefficient, and f represents the network's function. It is useful to note that the objective function can be changed based on the problem conditions. Cross-entropy and mean square error are the most popular objective functions. Finding the optimal value for each parameter requires investigating different combinations of values though this method is time-consuming and needs a considerable amount of memory for calculations. Gradient Descent (GD) is an applicable method based on finding the direction with a high negative slope to reach the minimum point. Therefore, all of the parameters are updated based on the changing of the objective function with new values to reach the minimal points, and the updated values can be defined as follows:

$$w_i + \alpha \frac{\partial J}{\partial w_i} \rightarrow w_{i+1} \tag{3-33}$$

where α is the learning rate, w_i is the value in this iteration, and w_{i+1} is the value in the new iteration. Use of the chain rule for differentiable functions is required for calculating $\frac{\partial J}{\partial w_i}$. To have a generalized view about this process, based on the random values for different weights and biases the training process is started. When the output data is calculated, according to the objective function the value of output is found. Then according to backpropagation method, the layers' coefficients from the last layer to first layer are updated. Afterwards, the process is repeated in the next iteration with updated amount. This process is continued until the optimal value is achieved for coefficients and objective function.

3.1.7 Deep Learning

Deep learning is a subset of machine learning that has a valuable ability to process and analysis big data [75]. Thus, the deep neural network with successive layers interprets each complex concept into more straightforward concepts in the output. Besides, it does not require full human supervision to determine the necessary information of the machine at any time. Like machine learning methods, deep learning is divided into supervised, unsupervised, semi-supervised, and reinforcement learning based on the data for analysing and the applications. The supervised deep learning scheme is illustrated in Figure 3-9.

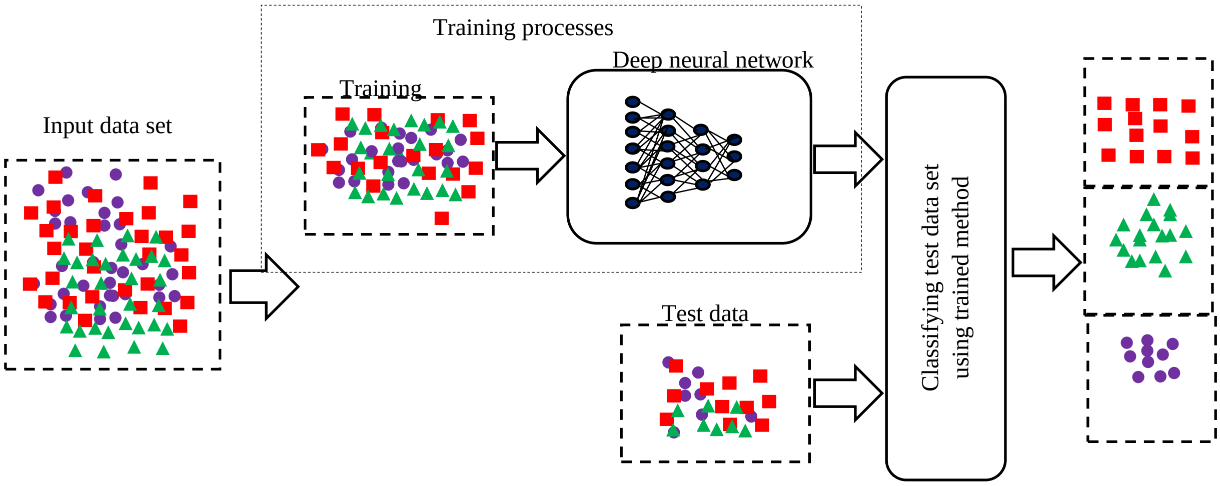


Figure 3-9: The supervised deep learning

One of the main important differences between deep learning and machine learning is pre-processing for extracting features as shown in Figure 3-10. Unlike the machine learning methods, the feature extraction process is done simultaneously with the training process in the deep learning method. However, using the feature extracting methods before entering the data into the network can be a valuable idea to enhance the network's performance. Besides, pre-processing methods can be a decent idea to compensate for the problems in finding the optimum amount for parameters in each layer.Deep learning is more effective for big data than other machine learning methods; using various layers in its structure with many neurons increases the training time and over-training chance. Therefore, using supercomputers with a powerful Graphics Processing Unit (GPU) and Central Processing Unit (CPU) substantially affects their performance.

3.2 Autoencoders

Autoencoders (AE) are unsupervised learning algorithms that reconstruct input data in the output layer [76]. Figure 3-11 shows that it is constructed from two symmetrical neural network: the encoder and the decoder. The encoder network maps the input data into the lower-dimensional space called hidden layer, and the decoder network reconstructs the higher-dimensional input data from the hidden layer’s data as output with a similar size as the input data. As it can be deduced from this structure, the network's goal is finding the best presentation of data in the hidden layer with reduced dimension for using as an input for reconstructing them in the output layer.

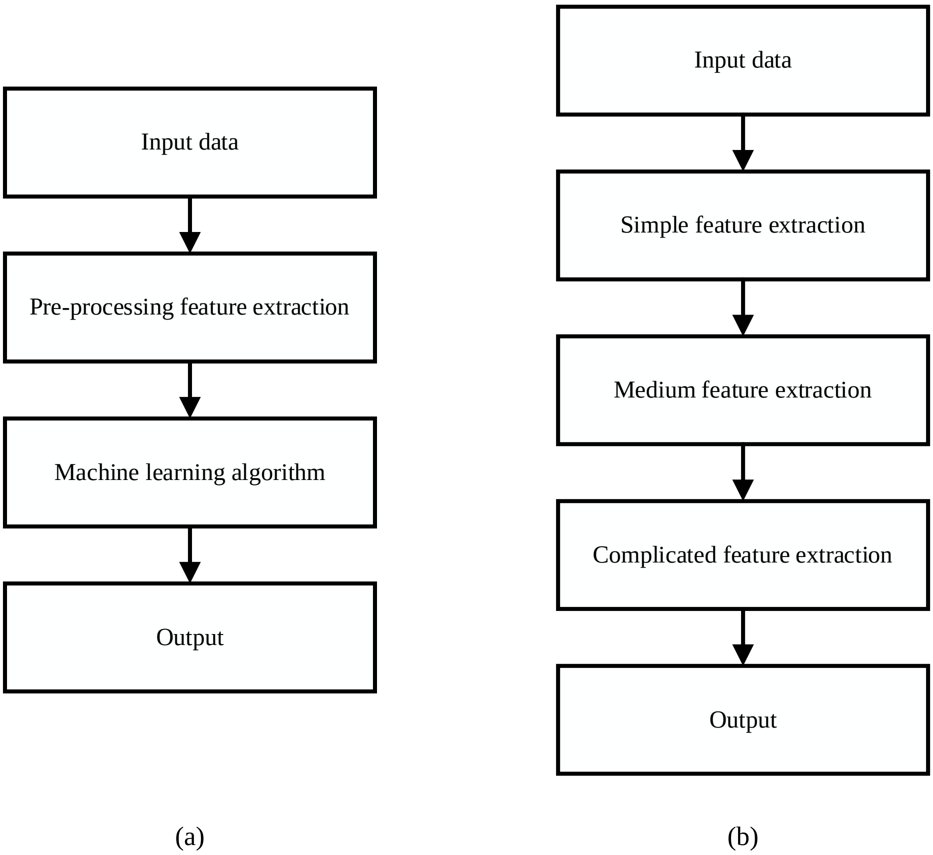


Figure 3-10: (a) Machine learning process (b) Deep learning process

According to the task of AE in reconstructing data, the loss function is defined simply as follows:

$$Loss = \|x - \hat{x}\|^2 \tag{3-34}$$

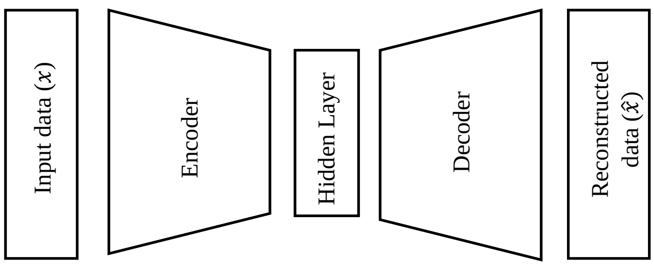


Figure 3-11: A simple schematic of Autoencoder

where x is the input data, and $\hat{x}$ is the output data. Denoising Autoencoder (DAE), Sparse Autoencoder (SAE), Contractive Autoencoder (CAE), and Variational Autoencoder (VAE) are different types of autoencoders each of which is prepared for different tasks. As it is found from DAE's name, the construction is based on reducing the noise from the data. In this method, the hidden layer is forced to learn robust features and prevent learning the identity function. It is crucial to prevent the neural network from learning the identity function as it promotes the acquisition of intricate and meaningful features that are pertinent to the given task.

The dimension of the hidden layer in a SAE is usually greater than the dimension of the input data. SAE employs a sparsity constraint in the objective function during training to induce a sparse representation of the data. This constraint encourages the network to activate only a small number of hidden neurons at a time, allowing SAE to learn more informative and robust features while also preventing overfitting. As a result, SAE can effectively capture complex and abstract features from the input data, which is beneficial for various applications such as speech and image recognition. CAE is another type of autoencoder that has an additional term in its loss function to be robust with respect to slight variation in input data. The difference between DAE and CAE is the definition of the objective function. In CAE's objective function, there is one term for penalizing the low-level variation in the input data. The VAE is another type of autoencoder with the same architecture compared to the autoencoder; however, this method uses a variational method as a learning method in its structure. In the following, this method is explained entirely.

3.2.1 Variational Autoencoder Method

A variational autoencoder is categorized as generative model in which the latent space are random variables characterized by probability distributions. The term "variable" comes from the relationship between regularity and the output of a statistical change. Generally, autoencoders cannot generate new data by sampling the latent space. The regularity of the encoding space of an autoencoder depends on the data distribution in the initial space before encoding, the dimensions of the encoding space, and the structure of the encoder. Therefore, it is difficult to guarantee that the encoder generates suitable data in the latent space. The network may be over-trained during training, or meaningless points in the latent space may get unwanted data when decoding. Thus, the latent space is prepared to be irregular, in which case there is some trouble getting new data, and it is not possible to choose with certainty from the hidden layer.

The basis of VAE is completely similar to that of an autoencoder in so far that they aim to minimize the reconstruction loss in order to reduce the difference between the original and the reconstructed data in the output. The latent space structure in a VAE is different from that of an AE. The latent variables of the VAE are the stochastic variables represented by the mean and variance of the distribution of data changed based on entering the new input data. It assumes that the latent variables follow a multivariate normal distribution with a diagonal covariance matrix. Additionally, the mean and variance are reasonable quantities to represent the occurring stochastic data because they provide a measure of central tendency and variability, respectively.

Therefore, the probability of reconstructing data with high accuracy in the VAE because of different parameters which control the distribution of data is higher than in an AE. The general process of VAE is presented as follows:

- The input data is initially encoded to the lower dimension space;
- The new data in the hidden space are generated based on gaussian distribution and the entrance of new data in the hidden space;
- The sample points are decoded, and the reconstruction loss is calculated;
- Train the network using the backpropagation method;

There is a comparison in Figure 3-13 between the structure of AE and the VAE. In practice, the distribution of the hidden layer is usually chosen as a normal distribution, and the encoder computes the mean and variance. The distribution generated by the encoder must be close to a normal distribution, and the latent space must be regular. Therefore, the objective function consists of a reconstruction term that enables data reconstruction and reduces loss and an adjustment term that approximates the distribution of the latent space to a normal distribution. This correction term is defined as the Kullback-Leibler (KL) difference between the constructed and standard Gaussian distributions. The objective function of VAE is presented as follows:

$$\text{Loss} = \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0,1)] \tag{3-35}$$

The hidden layer ought to have essential capabilities with a view to being organized and generating new data possible. VAE should have two important features to be successful in its performance: Continuity and Completeness. Continuity means that two close data points should not give two different contents in the new space. If the data points are mapped in the new space with a high distance, it deviates autoencoder and the reconstruction loss is increased. Besides, the data point in the new space according to the selected distribution should have a meaningful mapping with respect to the input. In the other words, the data before and after presence in the hidden layer should have the same label which this feature is known as completeness. The distributions are considered near to the Gaussian distribution, which covers nearly the complete hidden area. Usually, the value of doing that is to increase the reconstruction loss.

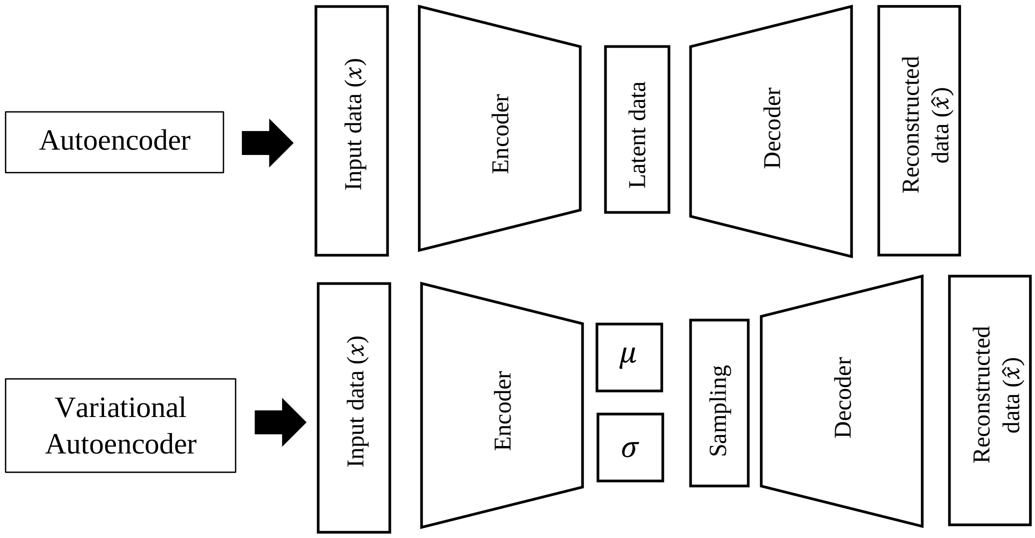


Figure 3-12: The autoencoder and variational autoencoder training process

3.2.1.1 Mathematics of Variational Autoencoder

The previous section gave a complete overview of how the VAE works. In the continuation of this section, the theory of VAE proposes and thoroughly explains regularization terminology. VAE is an autoencoder that encodes an input as a distribution rather than a point, and its potential spatial organization is adjusted by constraining the distribution returned by the encoder to be closer to the standard Gaussian distribution. The best way to tackle this stochastic framework is with variational reasoning. Assuming that x is a variable that represents the data and is generated from a latent variable z (a coded representation) that is not directly observed. Therefore, the following two-step generation process is assumed for each data point.

- First, the latent representation z is chosen from a prior distribution $\mathrm{p}(z)$;
- Second, the data x is chosen from a conditional likelihood distribution $\mathrm{p}(x|z)$;

Also, a stochastic decoder is naturally defined as $\mathrm{p}(x|z)$ that describes the distribution of the decoded variable under the hypothesis that the encoded variable possesses a certain distribution, and a stochastic encoder is defined with a $\mathrm{p}(z|x)$ that describes the distribution of the encoded variable. According to Bayes theorem:

$$\mathrm{p}(z|x) = \frac{\mathrm{p}(X|Z)\mathrm{p}(z)}{\mathrm{p}(x)} \tag{3-36}$$

It is assumed that the prior distribution $\mathrm{p}(z)$' follows a standard Gaussian distribution, while the conditional likelihood distribution $\mathrm{p}(x|z)$' is assumed to follow a Gaussian distribution with a positive constant 'c'. This distribution is averaged by the deterministic function 'f' of the variable 'z' and the covariance matrix is multiplied by the identity matrix 'I'. It is noted that 'f' is currently denoted by 'F' and its definition will be provided later. Given these assumptions, it is feasible to define the following:

$$\mathrm{p}(z) \equiv \mathcal{N}(0, I) \tag{3-37}$$

$$\mathrm{p}(x|z) \equiv \mathcal{N}(f(z), cI) \qquad f \in F \quad c > 0 \tag{3-38}$$

Variational inference (VI) is an efficient way to find a distribution equal to $\mathrm{p}(z|x)$. In VI, one distribution is approximated by a function to find the similarity between the objected distribution and the estimated distribution, where the estimated distribution can be a Gaussian distribution. It requires a mathematical tool to find the level of similarity, which can be done using a Kullback-Leibler (KL) divergence. Formally, the discrepancy KL between two distributions q and p with discrete support is defined as:

$$\mathrm{KL}(\mathrm{q}||\mathrm{p}) = -\sum_x \mathrm{q}(x) \log \frac{\mathrm{p}(x)}{\mathrm{q}(x)} \tag{3-39}$$

According to equation (3-39), three important conditional properties are deduced as follows:

- $\mathrm{KL}(\mathrm{q}||\mathrm{p}) \geq 0$ for all q,p
- $\mathrm{KL}(\mathrm{q}||\mathrm{p}) = 0$ when q=p
- $\mathrm{KL}(\mathrm{q}||\mathrm{p}) \neq \mathrm{KL}(\mathrm{p}||\mathrm{q})$

The purpose of this investigation is to evaluate the similarity between $\mathrm{p}(z|x)$ and $\mathrm{q}_x(z)$, which is a Gaussian distribution. The mean and variance of $\mathrm{q}_x(z)$are estimated using two functions, g (mean) and h (covariance). It is expected that these functions belong to the G and H function families, respectively. Therefore, it can be expressed as below:

$$\mathrm{q}_x(z) \equiv \mathcal{N}\big(g(x), h(x)\big) \qquad g \in G\,, h \in H \tag{3-40}$$

$g(x)$ and $h(x)$ are the goals to find the best one to interpret the approximation. In other words, g^* and h^* are found such that:

$$\begin{aligned}
&(g^*, h^*) = \underset{(g,h)\in G\times H}{\operatorname{argmin}}\ \mathrm{KL}\big(\mathrm{q_x}(z), \mathrm{p}(z|x)\big) = \\
&\underset{(g,h)\in G\times H}{\operatorname{argmin}}\ (\mathrm{E}_{z\sim x}(\log \mathrm{q}_x(z)) - \mathrm{E}_{z\sim x}(\log \tfrac{\mathrm{p}(X|Z)*\mathrm{p}(z)}{\mathrm{p}(x)})) = \\
&\underset{(g,h)\in G\times H}{\operatorname{argmin}}\ (\mathrm{E}_{z\sim x}(\log \mathrm{q}_x(z)) - \mathrm{E}_{z\sim x}(\log \mathrm{p}(z)) - \mathrm{E}_{z\sim x}(\log \mathrm{p}(x|z)) + \\
&\mathrm{E}_{z\sim x}(\log \mathrm{p}(x))) = \underset{(g,h)\in G\times H}{\operatorname{argmax}}\ (\mathrm{E}_{z\sim x}(\log \mathrm{q}_x(z)) - \mathrm{KL}(\mathrm{q}_x(z), \mathrm{p}(z))) = \\
&\underset{(g,h)\in G\times H}{\operatorname{argmax}}\ (\mathrm{E}_{z\sim x}(-\tfrac{\|x-f(z)\|^2}{2\mathrm{c}}) - \mathrm{KL}(\mathrm{q}_x(z), \mathrm{p}(z)))
\end{aligned} \tag{3-41}$$

According to equation (3-41), It can be seen that there is a compromise between maximizing the probability of observations and staying close to the previous probability (minimizing the divergence of KL), and a balance between the confidence in the data and the confidence in the total distribution. Also, to reduce the reconstruction loss, it is needed to have:

$$f^* = \underset{f\in F}{\operatorname{argmax}}(\ (E_{z\sim \mathrm{q}_x{}^*}(log\,\mathrm{p}(x|z)) = \underset{f\in F}{\operatorname{argmax}}(\ (E_{z\sim \mathrm{q}_x{}^*}\left(\left(-\frac{\|x-f(z)\|^2}{2c}\right)\right) \tag{3-42}$$

where $\mathrm{q}_x{}^*$depends on the function$f^*$and is obtained as described before. Gathering all the pieces together, f^*, g^* and h^* are found such that:

$$(f^*, g^*, h^*) = \underset{(f,g,h)\in F\times G\times H}{\operatorname{argmax}}\ (E_{z\sim x}(-\frac{\|x-f(z)\|^2}{2c}) - KL\big(\mathrm{q}_x(z), \mathrm{p}(z)\big)\big) \tag{3-43}$$

Equation (3-43) shows that the final objective function consists of two terms. The objective function seeks to decrease the reconstruction error in the initial term, while in the subsequent term, it aims to establish similarity between the $\mathrm{p}(z)$ and $\mathrm{q}_x(z)$ distributions, typically a Gaussian distribution. The neural network should be updated in each iteration, and according to the calculated error using the backpropagation method, the network parameters are regulated, and the sampling operation must be such as to enable this. According to the explanation of a VAE, it is clear when the network starts to regulate the coefficient using the backpropagation approach, it faces another problem. As the sampling process is a stochastic process, it is not possible to use derivatives in the process. Therefore, it is needed to use one trick to solve this problem in VAE. A simple trick called the reparameterization trick is used to enable gradient descent despite random sampling occurring in the middle of the architecture, where z is a random variable that follows a Gaussian distribution with mean $g(x)$ and covariance $h(x)$. Then it can be expressed as follows:

$$z = h(x)\zeta + g(x) \qquad \zeta \sim \mathcal{N}(0,1) \tag{3-44}$$

In the equation (3-44), the ζ is a gaussian distribution with zero mean and one variance. In each epoch, one random data is selected according to equation (3-44). As shown in Figure 3-13, when the network's training is implemented based on the traditional method, backpropagation is not possible. However, when the reparameterization trick is proposed, the network does not have any problem updating itself using the backpropagation method. According to previous research, it is sometimes better to use methods for reducing dimensionality to get better results. These methods can help the network to analyse faster and with higher accuracy. Besides, for implementing this method in different hardware with a lot of restrictions, it is better to find one method which reduces the input data, or it is better to say that planning data in different spaces which the reduced data in new space has the important feature of input data which can be replaced to the imported data.

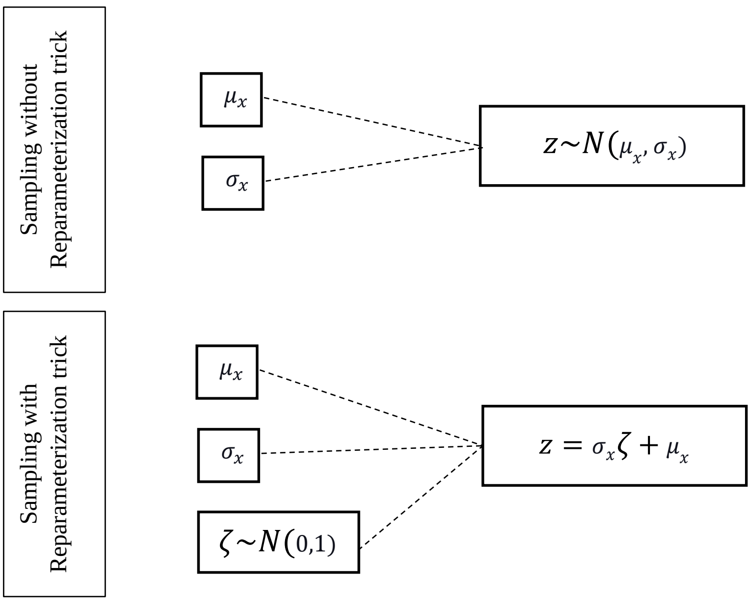


Figure 3-13: Reparameterization trick

It can be noted that the variational autoencoder method provides immediately a method for feature extraction. Consequently, it is not needed to use the other reducing dimensionality method. However, the two well-known methods to reduce dimensionality are described to get a better result in this research work. Feature extraction has many applications: face recognition, handwritten digit recognition, text mining, image retrieval, and fault detection; feature extraction is used for extracting essential features to get better separability.

3.2.1.2 The Process of Classifying using VAE

As described in Chapter 2, the goal of this thesis is the detection and classification of different faults; however, VAE is an unsupervised method whose primary task is reconstruction input data in the output. In this section, the process of using VAE for classifying data is defined. The process of preparing data for use as the input data in the VAE is shown in Figure 3-14. The most important step of dataset preparation is normalization which causes putting all data in the specified range. There are different methods for normalizing dataset and all of them is applicable. In this research work, all of the datasets are normalized using equation (3-45) as follows:

$$\text{Normalized value of } x_i = \frac{x_i - \min\text{(X)}}{\max(\text{X}) - \min\text{(X)}} \tag{3-45}$$

where x_i is the value of each sample in row i, and $\text{X} = \{\text{x}_1, \text{x}_2, \dots, x_i, \dots, \text{x}_\text{n}\}$. This process is done for all of the classes then they are organized in one file together. Shuffling dataset is the next step which causes to prevent overtraining in neural network. In the other words, this process causes datasets to have representatives in all batches and when training neural network is started the model is defined correctly.

Next, the dataset is divided to train and test datasets. 70% of the dataset belongs to train dataset and 30% to test dataset. Since it is not possible to enter all training and testing datasets into the network at once, the input size parameter is defined. Therefore, the data is divided into different sections based on input size parameter. The training of VAE and classifying datasets are described in the following.

3.2.1.2.1 Training Procedure

When the first section of train datasets is entered into the network, the training procedure starts, and the coefficients are updated according to the batch size parameter.

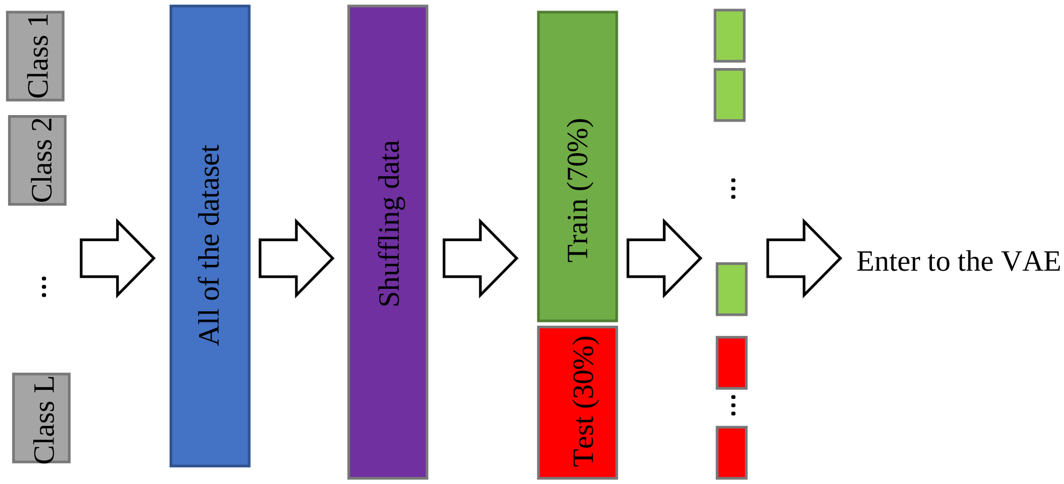


Figure 3-14: Data preparing diagram

Batch size is one of the most important parameters of each network for defining how much data is needed to enter to network and after that, the network's coefficients are updated. When the data is passed from the encoder, according to the reparameterization trick, new data is sampled based on the gaussian distribution which is considered in this research work. In the hidden layer, according to the dimension which is considered the new data is generated in different dimensions. Therefore, for each data that enter into the encoder, there is a vector in the hidden layer in the middle of the VAE to represent it in a reduced dimension. Next, the data in the hidden layer is sent to the decoder. Each data is reconstructed again in the output of the decoder. When the reconstructed datasets are defined completely, the reconstruction loss is calculated and the backpropagation approach starts to update the coefficients from the last layer of the decoder to the first layer of the encoder. It is important to note to this point that the ADAM optimizer is used in this process. The complete explanation of ADAM optimizer is presented in [77].

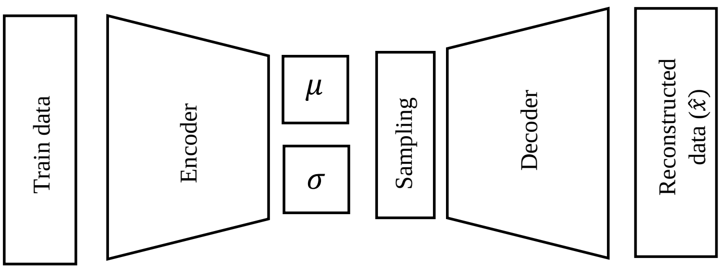


Figure 3-15: Training diagram of VAE

When all of the train datasets are sent to the network and the first iteration is finished, the next iteration is started for learning again. This process is continued based on the number of iterations called epochs, an important parameter of each network. Selecting the proper value for the epochs has an important effect on preventing overtraining in the network. The complete procedure of entering data until the training is finished shown in Figure 3-15.

3.2.1.2.2 Classification Procedure

When the training of the network for the first iteration is finished, the test datasets are entered into the network for investigating the performance of the network for classifying data as shown in Figure 3-16. The test datasets are passed from the encoder and the new data in the new space are generated similar to the training process.

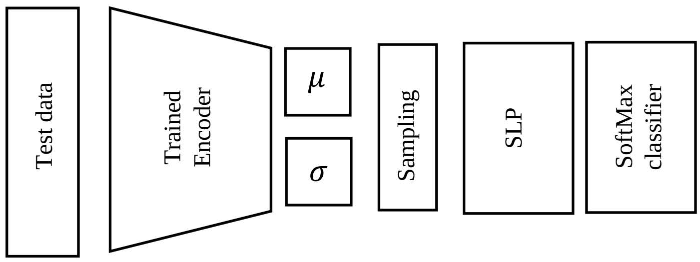


Figure 3-16: Classifying diagram for faults classification using VAE

In the continuation, instead of transferring data to the decoder for reconstruction, the data is sent to the classification section. In the classification section, there is one Single Layer Perceptron (SLP) network in which the number of neurons is the same as the number of classes. Therefore, there is an output layer similar size to the number of classes. Next, it is needed to use one classifier to show the class label. Therefore, the SoftMax classifier is used next to the output layer of the neural network. According to the explanation of SoftMax, the results of the output of each neuron is a value in the range of [0-1] and the summation of all of them is 1. Therefore, the highest amount among all of the neurons shows the label of data. When the label of each data is found, the result is compared to true labels. This process is done using one loss function called the cross-entropy loss function. The cross-entropy loss function is explained completely in [78]. Then based on the backpropagation approach the coefficients of SLP are regulated. A flowchart is prepared for describing the complete procedure of using VAE for training and classifying in Figure 3-17.

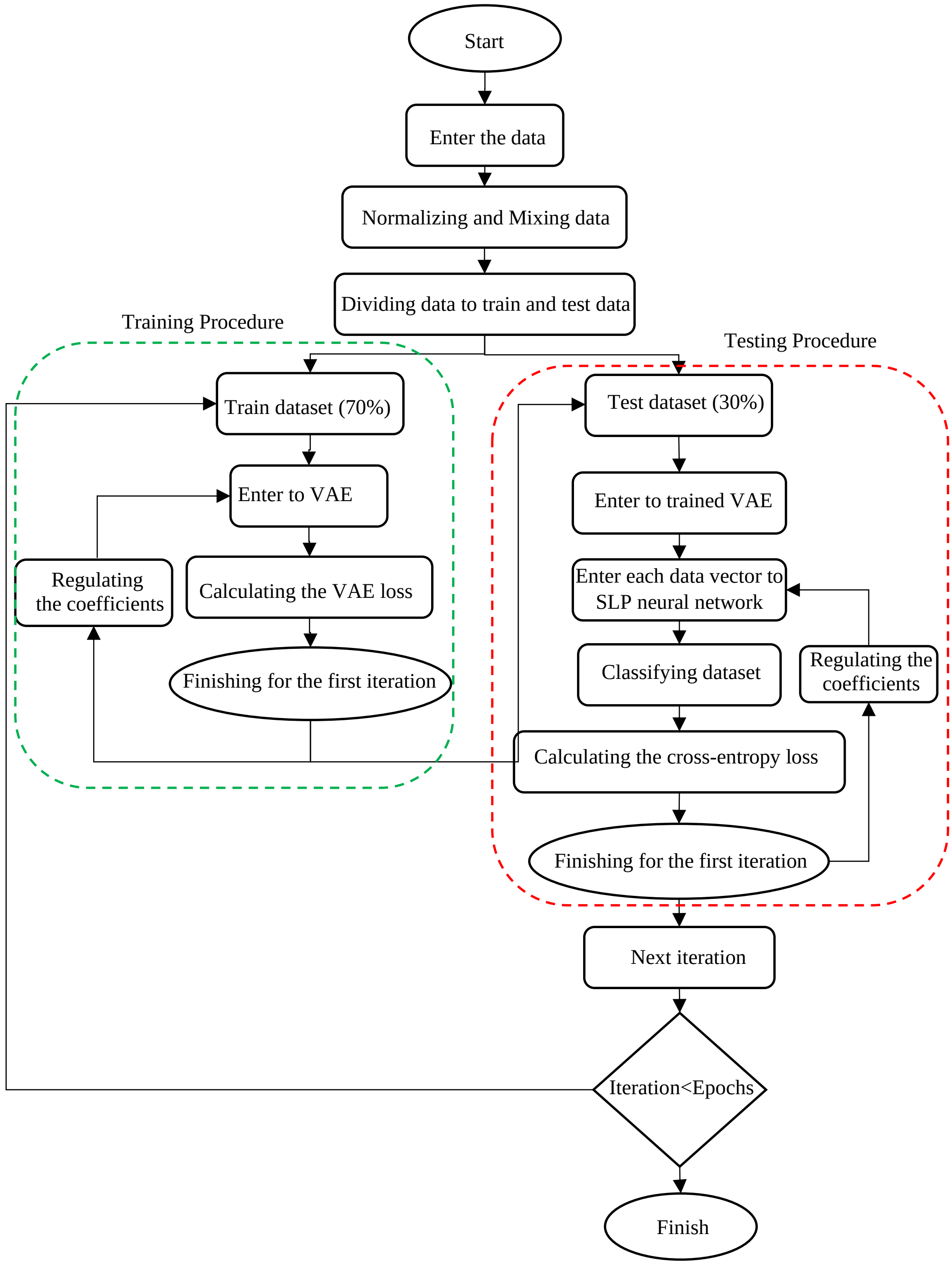


Figure 3-17: A general overview of using VAE for faults classification

4 EXPERIMENTAL AND SIMULATION SETUP

In this chapter, the different components of the experimental setup are investigated entirely. This chapter consists of four basic sections. First, valuable information about designing and the mathematical model of PMSM is presented. Then, the simulated PMSM using Finite Element Analysis (FEA) is compared to the prototyped one to investigate the difference between the simulation and the prototyped machine. Next, helpful information about implementing various faults in the machine are prepared. Finally, valuable information about the sensors and data acquisition card are given.

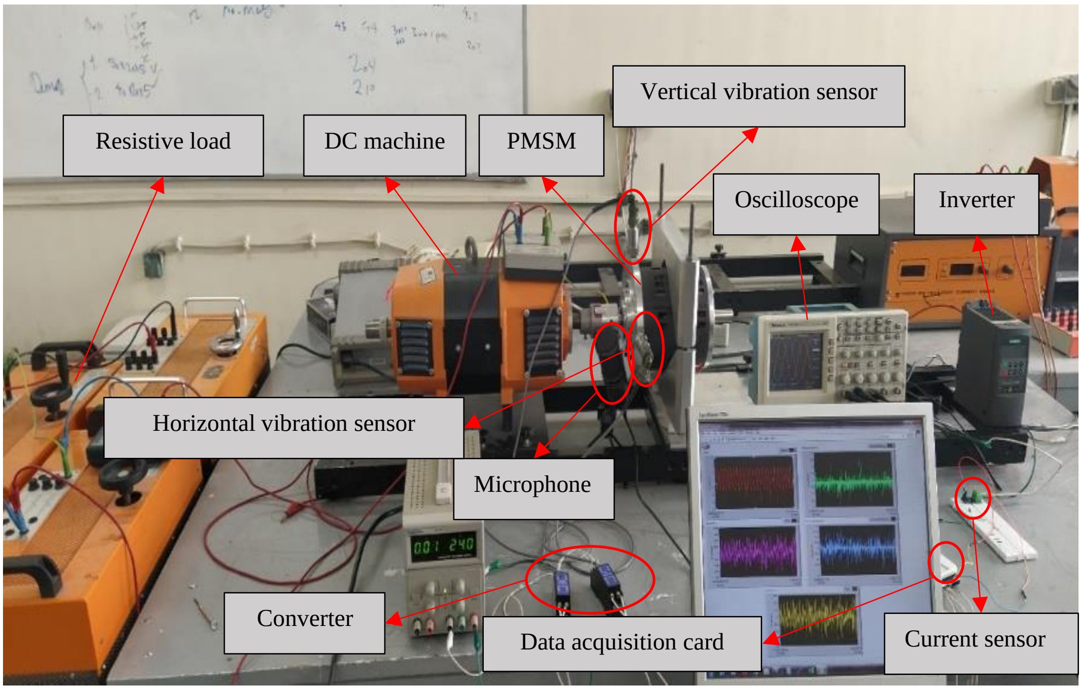


Figure 4-1: Experimental setup

4.1 Experimental Setup

Figure 4-1 illustrates the experimental setup for gathering data from PMSM for various states. The critical part of this experimental setup is two electrical machines: a permanent magnet synchronous and a Direct Current (DC) machine. The PMSM is coupled to the DC machine working as a generator connected to resistive load. In this project, a 1 kW PMSM is designed, simulated, and prototyped to implement different defects with different severity. The PMSM is powered by a 3 kW electric drive of inverter model Micromaster 420 manufactured by Siemens, as shown in Figure 4-1. The various levels of loads is implemented on the PMSM shaft by changing the resistance load which is connected to the armature winding of the DC machine. In this project, two vibration sensors model SLD 144B manufactured by SPM company are used to measure vibration in the vertical and horizontal directions. A current sensor model ACS712 manufactured by Allegro is used to collect the current of winding phase A online.

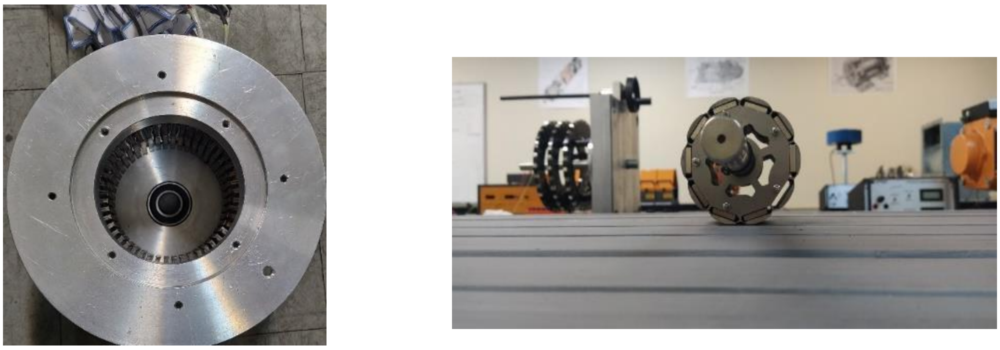
Table 4-1: PMSM design information

Parameters	value
Rated power	1 kW
Rated voltage	230 V
Rated frequency	125 Hz
Rated speed	1500 RPM
Rated torque	6.36 Nm
Inside radius of the stator core	47.5 mm
Length of the stator core	57.5 mm
Air gap length	0.5 mm
Type of bearing	6204 2RS C3TN
Magnet model	N42
B_r	1.293 T
Magnet's mass density	7500 $^{kg}/_{m^3}$
Magnet's dimension (length-width-height)	50 mm-20 mm-5 mm
Magnet's position	Buried magnets
Pole pairs number	5
Width of stator slot	23.45 mm
Height of stator slot	5.1 mm
Number of slots	48

Besides, a microphone for collecting sound signals manufactured by Audio-Technica, is used. The data are collected using a national instrument data acquisition card model USB-6009 and transferred to LabVIEW as interface software. The process of data acquisition is described in the following. The inverter produces the needed voltage for rotating PMSM coupled to the DC machine. Therefore, the PMSM is coupled to the DC machine, which works as a generator, the field winding connected to dc voltage, and the armature winding connected to resistance load. When the PMSM rotates the DC machine, the current starts to pass from the armature winding, and the armature reaction is produced, which causes a load on the shaft of the PMSM. Increasing and decreasing resistive load changes the load torque of the machine. Then the data is gathered using sensors and transferred to the data acquisition card.

4.2 Permanent Magnet Synchronous Motor

It is important point to note that designing, optimizing, and building process of this electrical machine are not the content of this thesis. Nevertheless, it is better to clarify some essential components of the electric machine. The design information of the PMSM is presented in Table 4-1.



(a) (b)

Figure 4-2: A view of PMSM components (a) Stator (b) Rotor

The stator and rotor of the machine after disassembling for implementing faults are illustrated in Figure 4-2. When selecting a winding for an electric machine, it is important to consider several crucial features to ensure its efficient and reliable performance. The winding should be properly insulated to prevent any short-circuiting or damage to the machine structure. It should also be designed uniformly with consistent wire diameter, number of turns, and turn spacing to ensure an even flow of current, minimizing the risk of hot spots or uneven wear. Additionally, the winding should be able to withstand the mechanical and thermal stresses that occur during operation and designed to be efficient to minimize power losses. The winding coefficient, low leakage inductance, and high fill factor should also be optimized to utilize the magnetic field effectively. In addition to these features, the length of the end winding should be minimized, and the placement of the winding should be optimized. The windings are divided into two categories: concentrated and distributed windings. The coils wound around each tooth in the concentrated winding, and the number of slots per pole per phase (q) is fractional and lower than 1. In contrast, q has different values based on different combinations of slots and poles for the distributed winding.

On one hand, the distributed winding increases the copper losses and the weight and volume of the prototyped machine. On the other hand, concentrated winding possesses features such as a short end winding, simpler structure, and a more reliable insulation system. The drawback of concentrated windings compared to distributed windings is its production of higher harmonics. In low-speed machines, the concentrated winding is more efficient, while in high-speed machines, distributed windings are used. Also, another important parameter is the winding configuration. Generally, there are full-pitch and fractional windings. In full pitch winding, the coil pitch is equal to the pole pitch, while in fractional or short pitch winding, the coil pitch is shorter than the pole pitch. The main problem with short pitch winding is the reduction of the winding coefficient to reach the sinusoidal voltage waveform. In this research work, the distributed winding is selected to be implemented. The dimensions of stator and rotor yoke are determined based on different parameters such as the pole number, air gap flux, and permissible flux in the core. The mathematical model of PMSM is described in the following.

4.2.1 Mathematical Model of PMSM

As explained in the previous chapters, the PMSM is an electrical machine that uses PMs in its structure. They produce the needed Magnetic-Motive-Force (MMF) of the machine in the interaction of other parts of the machine’s winding. This feature let to not use a lot of winding in the structure causing to reduce the weight and volume of the machine. Besides, the motor's efficiency is increased because of reducing the winding loss, and the motor can be operated at high torque. PMSMs can generally be classified into different types of machines based on the configurations of the rotor and stator poles as well as the direction of the generated flux; however, the two main applications of these machines are axial and radial flux machines. In the radial flux machine, the direction of the air gap flux is along the radius of the machine and in axial flux, it is in along the shaft of the rotor. As shown in Figure 4-3, according to magnets location in the rotor, they are categorized to three types of machines; (a) Surface Mounted PMSM (SPMSM), (b) Interior PMSM (IPMSM), and (c) Surface Inset Mounted PMSM (SIPMSM). In the SPMSM which the magnets are located on the surface of rotor, the magnetic flux of magnets is high because of the magnets are in contact with air directly. However, this type of machines is not suitable for high speed because of lower structure robustness.

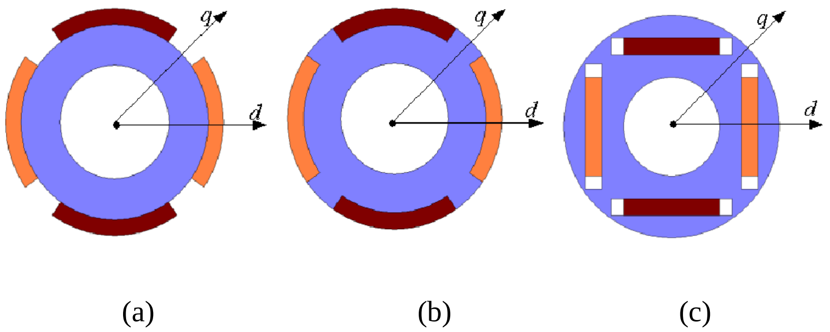


Figure 4-3: Different configuration of PMSM based on magnets placement
(a) SPMSM (b) Surface Inset Mounted PMSM (c) Interior PMSM [79]

In contrast, the IPMSM and SIPMSM have magnets positioned within the rotor core. Therefore, they are proper for high speed machines. It can be noted that SPMSM type is used in this research work. The cross-sectional schematic of a two-pole, three-phase PM synchronous motor is shown in Figure 4-4. The mathematical model of electrical machines for three phases can be defined as follows [80]:

$$V_{\mathrm{a}} = R_{\mathrm{a}} i_{\mathrm{a}} + L_{\mathrm{a}} \frac{\mathrm{d}i_{\mathrm{a}}}{\mathrm{d}t} + M_{\mathrm{ab}} \frac{\mathrm{d}i_{\mathrm{b}}}{\mathrm{d}t} + M_{\mathrm{ac}} \frac{\mathrm{d}i_{\mathrm{c}}}{\mathrm{d}t} + \omega_{\mathrm{e}} \lambda_{\mathrm{a}} \tag{4-1}$$

$$V_{\mathrm{b}} = R_{\mathrm{b}} i_{\mathrm{b}} + L_{\mathrm{b}} \frac{\mathrm{d}i_{\mathrm{b}}}{\mathrm{d}t} + M_{\mathrm{ba}} \frac{\mathrm{d}i_{\mathrm{a}}}{\mathrm{d}t} + M_{\mathrm{bc}} \frac{\mathrm{d}i_{\mathrm{c}}}{\mathrm{d}t} + \omega_{\mathrm{e}} \lambda_{\mathrm{b}} \tag{4-2}$$

$$V_{\mathrm{c}} = R_{\mathrm{c}} i_{\mathrm{c}} + L_{\mathrm{c}} \frac{\mathrm{d}i_{\mathrm{c}}}{\mathrm{d}t} + M_{\mathrm{ca}} \frac{\mathrm{d}i_{\mathrm{a}}}{\mathrm{d}t} + M_{\mathrm{cb}} \frac{\mathrm{d}i_{\mathrm{b}}}{\mathrm{d}t} + \omega_{\mathrm{e}} \lambda_{\mathrm{c}} \tag{4-3}$$

where V_{a}, V_{b}, and V_{c} are the voltage of phases a, b and c, respectively. Also, R_{a}, R_{b}, and R_{c} are three-phase resistances which in the ideal situations are equal. The self-inductances for different phases are L_{a}, L_{b}, L_{c}, and M_{ij} are mutual inductances between phase i and j. Besides, i_{a}, i_{b}, and i_{c} are three-phase currents, and ω_e is the electrical angular speed. The three-phase flux linkages for three-phases are λ_{a}, λ_{b}, λ_{c}.

The flux linkage of each phase produced by magnets can be written as follows:

$$\lambda_{\mathrm{a}} = \lambda_{\mathrm{pm}} \sin(\theta) \tag{4-4}$$

$$\lambda_{\mathrm{b}} = \lambda_{\mathrm{pm}} \sin(\theta - 2\frac{\pi}{3}) \tag{4-5}$$

$$\lambda_{\mathrm{c}} = \lambda_{\mathrm{pm}} \sin(\theta + 2\frac{\pi}{3}) \tag{4-6}$$

where λ_{pm} is the PM flux linkage and θ is angle between rotor and stator. If the equations (4-4)-(4-6) are replaced in equations (4-1)-(4-3), the complexity of these formulations is increased.

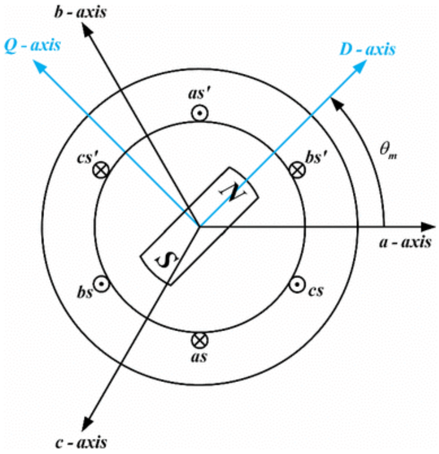


Figure 4-4: Cross-sectional schematic of a two-pole, three-phase PM synchronous motor [81]

The best way to reduce the complexity is to transform the aforementioned equations from the stator to rotor frames. As the equations are presented in the rotor frame the dependency on θ in equations (4-4)-(4-6) is removed. To achieve this goal, first the equations are transformed to a stationary frame $(\alpha - \beta)$ plane using the Clarke transformation. Then they are transformed to the rotor frame by another transformation. The Clarke transform in this case is defined as follows [80]:

$$x_{\alpha\beta\gamma} = \mathrm{T_a}.x_{\mathrm{abc}} \tag{4-7}$$

$$\mathrm{T_a} = \sqrt{\frac{2}{3}}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{4-8}$$

Also, it is possible to use Park's transform to transform from stator frame (abc) to rotor frame (dq0) directly to reduce the complexity as follows [80]:

$$x_{\mathrm{dq0}} = \mathrm{T_P}.x_{\mathrm{abc}} \tag{4-9}$$

$$\mathrm{T_P} = \begin{bmatrix} \cos(\theta) & \cos(\theta - 2\frac{\pi}{3}) & \cos(\theta + 2\frac{\pi}{3}) \\ \sin(\theta) & \sin(\theta - 2\frac{\pi}{3}) & \sin(\theta + 2\frac{\pi}{3}) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{4-10}$$

where in the rotor frame, d-axis (direct axis) is the direction of generating of flux in the machine, and q-axis (quadrature axis) is 90 electrical degree after d-axis. Besides, it is possible to transform from dq0 to abc frame again using follows equations:

$$x_{\mathrm{abc}} = \mathrm{T_P}^{-1}.\, x_{\mathrm{dq0}} \tag{4-9}$$

$$\mathrm{T_P}^{-1} = \frac{2}{3}\begin{bmatrix} \cos(\theta) & \sin(\theta) & 1 \\ \cos(\theta - 2\frac{\pi}{3}) & \sin(\theta - 2\frac{\pi}{3}) & 1 \\ \cos(\theta + 2\frac{\pi}{3}) & \sin(\theta + 2\frac{\pi}{3}) & 1 \end{bmatrix} \tag{4-11}$$

Therefore, the mathematical equations of machine transform from stator frame to rotor frame reads as follows:

$$V_\mathrm{d} = r_\mathrm{s} i_\mathrm{d} + L_\mathrm{d}\frac{\mathrm{d}i_\mathrm{d}}{\mathrm{d}t} - \omega_\mathrm{e} L_\mathrm{q} i_\mathrm{q} = r_\mathrm{s} i_\mathrm{d} + \frac{\mathrm{d}\lambda_\mathrm{d}}{\mathrm{d}t} - \omega_\mathrm{e}\lambda_\mathrm{q} \tag{4-12}$$

$$V_\mathrm{q} = r_\mathrm{s} i_\mathrm{q} + L_\mathrm{q}\frac{\mathrm{d}i_\mathrm{q}}{\mathrm{d}t} + \omega_\mathrm{e} L_\mathrm{d} i_\mathrm{d} + \omega_\mathrm{e}\lambda_\mathrm{pm} = r_\mathrm{s} i_\mathrm{q} + \frac{\mathrm{d}\lambda_\mathrm{q}}{\mathrm{d}t} + \omega_\mathrm{e}\lambda_\mathrm{d} \tag{4-13}$$

$$\lambda_\mathrm{d} = L_\mathrm{d} i_\mathrm{d} + \lambda_\mathrm{pm}, \qquad \lambda_\mathrm{q} = L_\mathrm{q} i_\mathrm{q} \tag{4-14}$$

where V_d, V_q are direct and quadrature axis voltages, i_d, i_q are direct and quadrature axis currents, r_s is the stator resistance, L_d and L_q are direct and quadrature axis inductances, λ_d and λ_q are direct and quadrature axis flux linkages. Besides, when the motor operates in the steady state while the torque of the machine is constant, derivative parts of equations (4-12)–(4-13) are removed and the equations change as follows in steady state:

$$V_\mathrm{d} = r_\mathrm{s} i_\mathrm{d} - \omega_\mathrm{e}\lambda_\mathrm{q} \tag{4-15}$$

$$V_\mathrm{q} = r_\mathrm{s} i_\mathrm{q} + \omega_\mathrm{e}\lambda_\mathrm{d} \tag{4-16}$$

also, the generated torque of PMSM is presented as follows:

$$\mathrm{T} = \frac{3p}{2}\left(\lambda_\mathrm{d} i_q - \lambda_\mathrm{q} i_\mathrm{d}\right) = \frac{3p}{2}\left(\lambda_\mathrm{pm} i_q - (L_\mathrm{d} - L_\mathrm{q}) i_\mathrm{d} i_\mathrm{q}\right) \tag{4-17}$$

where p is the number of pole pairs.

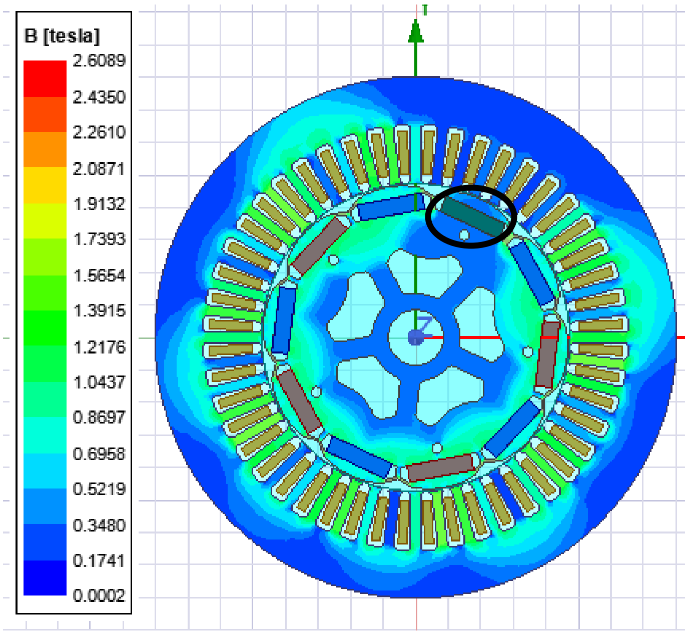


Figure 4-5: Simulation of PMSM in Finite element analysis with demagnetization

4.2.2 Simulation of PMSM

This section investigates the simulated PMSM in the Finite Element Analysis (FEA) software and the prototyped PMSM. This analysis is done in healthy and faulty states. The specified magnet in Figure 4-5, is replaced with non-magnetic steel to simulate partial demagnetization in PMSM. The simulation of magnetic field distribution is illustrated in Figure 4-5. This distribution is not uniform and makes an unbalanced magnetic pull in the machine.

Figure 4-6 illustrates the difference between the extracted EMF signals from the simulation and the experimental machine. Figure 4-6 (a) and (c) show that demagnetization with 100% severity reduces the EMF signal's peak magnitude from 150 V to 125 V. Besides, Figures 4-6 (b) and (d) show the EMF signal's frequency domain for the healthy and faulty situations. There is a minor difference between the extracted signal of the prototyped and simulated machine for the rated condition. As shown in Figure 4-6 (d), the frequency spectrum of the prototyped and the simulated machine shows low manufacturing tolerance. In conclusion, FEA simulation can be useful for analyzing similar to the prototyped machine; however, in the continuation of this research work, all the analysis is done based on the extracted signals from the prototyped PMSM.

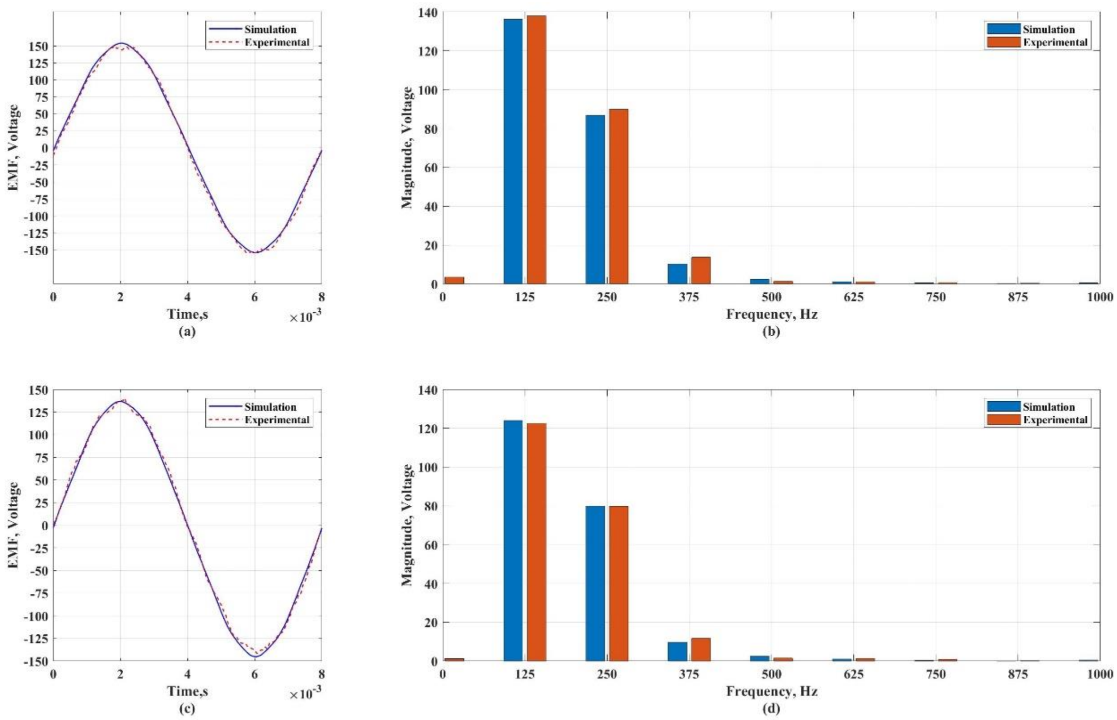


Figure 4-6: Comparison between simulation and experimental results for normal and one demagnetized magnet. (a) real time for normal; (b) Frequency spectrum for normal; (c) real time for one magnet demagnetization; (d) Frequency spectrum for one magnet demagnetization

4.3 Implementation of Faults in PMSM

This section presents a detailed description to implement different faults in the PMSM. As described in the previous sections, four different types of faults are considered to be implemented on a PMSM; demagnetization, bearing, eccentricity, and short-circuit fault. However, collecting data for short-circuit fault needs more time for implementation. In addition, short-circuit fault can damage machine windings and affects the other's faults' implementation. Therefore, the data are collected for demagnetization, bearing, and eccentricity with various severity in different operating conditions.

4.3.1 Implementation of Demagnetization

The fundamental goal of this thesis is the diagnosis of demagnetization in PMSM. As described, demagnetization is a magnetic fault of PM machines that causes a non-uniform magnetic distribution within the air gap and an unstable dynamic of the machine. Therefore, it is essential to implement the fault accurately. According to the review of demagnetization detection methods in Chapter 2, there are several methods to implement demagnetization, such as using an oven or removing one segment of the magnet to reduce the magnetic property of the magnet. The best way to implement demagnetization is to increase the temperature by using the oven to lose some magnetic properties. The benefits of this method are controllable fault implementation. However, due to the lack of access to such an oven, this technique was not possible. Therefore, implementing magnet cutting is another choice. However, cutting the magnet using the Wire-Cutter instrument might make the magnet lose its magnetic characteristic partially or completely. Thus, this method is not a controllable method, and it is better to find another method to get a better result. In the continuation, the healthy magnet is replaced by steel-304 with 7930 $({}^{kg}/_{m^3})$ mass density in various sizes for simulating different severity.

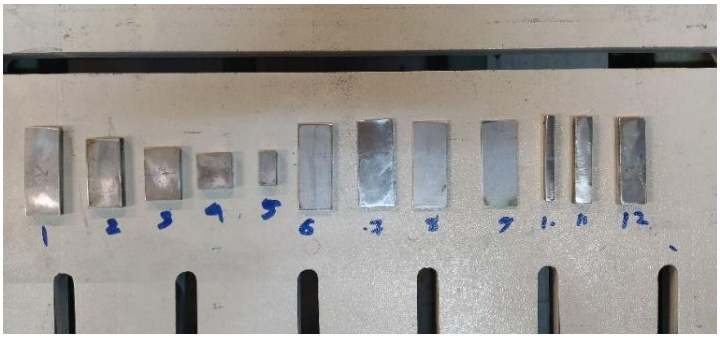


Figure 4-7: Non-magnetic pieces for simulating demagnetization

The volume space of the slot is filled with the combination of a lower-sized PM and steel. For instance, 30% demagnetization is simulated by replacing a healthy PM with a combination of 30% PM and 70% non-magnetic steel. Figure 4-7 illustrates the various sizes of steel for simulating different severity of demagnetization. Using non-magnetic steel makes accurate simulation and has an important effect on maintaining dynamic stability in the rotation of PMSM. Therefore, steel-304 is used in different sizes for substituting demagnetized parts of magnets. Also, using different PMs helps to implement fault in different directions. Table 4-2 illustrates the different sizes of steel-304 for implementing various severity of demagnetization in PMSM. In the continuation of this research work, demagnetization with different severity is applied in different directions because of the importance of detecting demagnetization in different directions. As seen, scenarios No.0 and 1 are considered to simulate full demagnetization in various situations with and without steels.

Table 4-2: Different size of steel-304 for simulating demagnetization

Fault	Direction of Fault	Scenario No.	Steel Dimension	Severity in Percent (%)
Demagnetization	All	0	No-steel	100%
		1	50-20-5	
	Z-axis	2	50-20-1	20%
		3	50-20-2	40%
		4	50-20-3	60%
		5	50-20-4	80%
	Y-axis	6	50-5-5	25%
		7	50-10-5	50%
		8	50-15-5	75%
	X-axis	9	10-20-5	20%
		10	20-20-5	40%
		11	30-20-5	60%
		12	40-20-5	80%

4.3.2 Implementation of Bearing Faults

Bearing fault is one of the prevalent faults in all electrical machines. Also, it can happen in all components of the bearing, consisting of the ball, inner ring, outer ring, and cage. Disassembling in most ball-bearing is not possible for implementing faults. In this research, the ball-bearing model 6204 2RS C3 TN is used. The model of ball-bearing is a commonly used type of bearing in various industrial applications. It consists of an inner ring, an outer ring, and a set of metal balls that are placed between them. The bearing is designed to reduce friction and support the load of rotating machinery, such as motors and gearboxes. The 6204 in the model number indicates the size of the bearing, while 2RS denotes that it has two rubber seals to protect against contamination. C3 represents the internal clearance, which is the amount of space between the balls and the races, and TN stands for a high-speed nylon cage that holds the balls in place. The ball-bearing model 6204 2RS C3 TN is known for its durability and reliability, making it a popular choice.

(a)

(b)

(c)

(d)

Figure 4-8: Implementing bearing faults on ball bearing. (a) Inner ring with different severity of faults; (b) Outer ring with different severity of faults; (c) Ball with different severity of faults; (d) The used ball bearing with model name

According to the characteristics of this type of bearing, it is a single row deep groove ball bearing with an inner diameter of 20 mm with double side rubber sealed with a polyamide cage. The polyamide cage causes to disassemble all of the bearing components easily; however, it is not possible to apply cage fault in this type of bearings. Therefore, the cage fault is not considered on this bearing because of using polyamide material for its cage. Bearing faults are applied to bearing components by holing on different bearing components with different sizes using Electro Discharge Machining (EDM).

Table 4-3: Different size of steel-304 for simulating bearing faults

Fault		Hole size in mm (width, depth)				
Bearing	Inner Ring	1, 0.1	1.25, 0.1	1.5, 0.1	1.75, 0.1	2, 0.15
	Outer Ring	1, 0.1	1.25, 0.1	1.5, 0.1	1.75, 0.1	2, 0.15
	Ball	1, 0.1	1.25, 0.1	1.5, 0.1	1.75, 0.1	2, 0.15

Also, the hole sizes are defined based on the precision of available EDM tools. It is important to note the tolerance of applying faults to have a precise fault severity. According to declaring of the EDM operator, holing lower than 1 mm in width and 0.1 mm in depth is not possible in this case study. However, there is this possibility that other EDM tools to apply lower fault severity precisely. Figure 4-8 depicts the different bearing components with various hole sizes for simulating bearing faults of different severity. Also, the different considered scenarios of bearing faults are presented in Table 4-3.

4.3.3 Implementation of Eccentricity

Eccentricity is one of the faults in which the harmonic features are the same as the demagnetization fault in some cases. Therefore, separating demagnetization from eccentricity fault is hard. As described in Chapter 2, eccentricity has the three types static, dynamic, and mixed. However, it is not possible to investigate all kinds of this fault in this research work, and most of the concentration is on the static fault. This fault is implemented by changing the air gap distance on one side of PMSM using different caps with the different centric points on one side of the machine, as shown in Figure 4-9. Different severity of eccentricity is implemented with varying eccentric lengths on one side of machine, making non-uniformity on the airgap. Applying eccentricity lower than 0.1 mm can increase the possibility of collision of the rotor and stator. Four types of eccentricity in this machine are considered consisting of 0.1mm, 0.15mm, 0.2mm, and 0.25 mm, which are labelled as Ecc-01, Ecc-02, Ecc-03, Ecc-04, respectively.

Figure 4-9: Implementing different eccentricity faults

4.4 Data Acquisition Setup

The critical part of the experimental setup is the data acquisition system. It can be helpful to describe the general data collection process in this experimental setup. The data from the vibration sensors go to the converter, which transforms the vibration values to voltage in the 0-5 V range. Also, the current signal is gathered using a connection coming from the inverter output to the sensor, and the output of the sensor goes to the machine terminals. Besides, the microphone directly connects to the computer to collect the sound signals via a USB connection. Finally, all data goes to the data acquisition card and are synchronized in the LabVIEW software. All of the data is extracted with a 16 kHz frequency rate. The sensors and data acquisition card used are introduced in the following. Besides, it is essential to note that a proper type of wire was used in this research work, which had an electromagnetic shield for protecting data against noise.

4.4.1 Sensors

According to the data collection aims, proper sensors are selected consisting of voltage, current, vibration, and sound sensors. Besides, the voltage sensor is used to gather the EMF of the machine for comparing prototyped and simulated machines. In the following, all of the sensors are explained entirely.

4.4.1.1 Voltage Sensor

As described in pervious sections, the voltage signal is gathered using the voltage sensor model ZMPTB101b, shown in Figure 4-10. According to the data sheet of this sensor, the capability of this sensor to gather the voltage is 250V AC. Therefore, it can be proper for collecting voltage from the prototyped machine because the rated voltage is 230V AC. It can be noted that this type of sensor, which can be seen in Figure 4-10. is a transformer that converts the AC voltage 0-250V to 0-5V in the output of the sensor, and it is acceptable for the input of the data acquisition card. The input of the data acquisition card in most cases is 0-20mA or 0-5V. When the input voltage is zero, the output shows a 2.5V offset. Besides, it is crucial to understand that the output of the current sensor goes to the data acquisition card, which is described in the following.

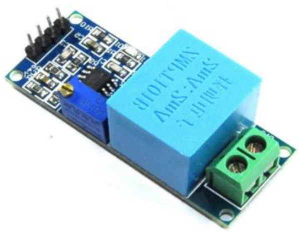


Figure 4-10: The voltage sensor model ZMPT101B

4.4.1.2 Current Sensor

One of the most important signals is current, which in recent research, much progress has done on it. The Motor Current Signature Analysis (MCSA) is one of the most popular methods to gather current signals and processes with different time-frequency analyses such as FFT.

Figure 4-11: The current sensor model ACS712

In this research, a hall effect current sensor model ACS712 manufactured by Allegro company is used, as shown in Figure 4-11. This sensor's sensitivity is 185mV/A, and it can measure until 5A. According to the low rated current of the prototyped machine, the 5A model is selected. Besides, the output of this sensor is 0-5V, similar to the voltage sensor. When the current is zero, there is an offset of 2.5V in output.

4.4.1.3 Vibration Sensor

One of the most effective ways to monitor industrial electrical machines' health levels is using a vibration sensor. When the fault occurs, the machine's vibration increases in various directions based on the fault severity and location. Therefore, detecting faults using vibration sensors can be an effective method to diagnose the faults. In this research, the vibration transducers manufactured by SPM company model SLD144B are used, as shown in Figure 4-12 (a). These transducers cannot be directly connected to the data acquisition card.

As shown in Figure 4-1, the output of vibration transducer should be converted in range of 0-5V and denoised using the vibration converter. According to its datasheet, the sensitivity is 100 mV/g in the frequency range of 2 Hz-10 kHz. In this project two vibration sensors are used in vertical and horizontal directions, as shown in Figure 4-12 (b).

(a) (b)

Figure 4-12: The vibration transducer model SLD144B (a) View of sensor, (b) Position of sensors on PMSM

4.4.1.4 Sound Sensor

One of this project's essential goals is to find the cheapest way to detect faults. According to the literature review done in Chapter 2, a microphone can be a valuable tool to diagnose the fault inexpensively with high accuracy. Figure 4-13 shows the AT2020USB manufactured by Audio-Technica used in this research work. This microphone is one of the cheapest microphone models on the market. It has two channels with a 48 kHz sampling rate. Besides, an analogue to digital converter exists with the microphone and is not needed to use another converter. Also, it can be connected to the computer via a USB connector. This microphone is a cardioid type that amplifies the sound in front and reduces the effects of environmental noises. Also, according to the frequency capability diagram of this microphone, it works in the frequency range of 20 Hz-20 kHz.

Figure 4-13: The microphone model AT2020USB

4.4.2 Data Acquisition Card

The essential part of the experimental setup is the data acquisition card. Figure 4-14 shows the USB-6009 manufactured by National Instrument, a multifunctional device with various analogue/digital inputs and outputs for gathering and generating signals.

This type of card is light and small, and has two analogue outputs with a 150 Hz frequency rate to generate signals and eight analogue inputs with 14-bit resolution and 48 kHz sampling frequency non-simultaneously. For instance, each channel can sample with a 16 kHz frequency rate when three analogue inputs are used. This card is connected to the computer via a USB connector, and data are collected using LabVIEW as interface software. In the summary of this chapter, the different component of the experimental setup for this research work was described thoroughly, and the extracted EMF of the prototyped and simulated machine was compared. Also, it was revealed that the prototyped and simulated machine results are approximately the same. Besides, the mathematical model of the machine was described, and the important equations of the machine were explained. In the next chapter, the extracted results of implementing VAE on extracted signals are analysed.

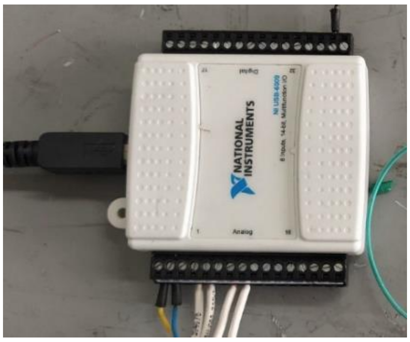


Figure 4-14: The data acquisition card model USB6009

5 DEVELOPMENT OF THE DETECTION METHOD

This chapter provides the performance results of using VAE as a new method for detecting demagnetization and other faults in various loads and speeds. According to the goals of this thesis described in Chapter 2, this method should be benchmarked from different aspects and compared with different machine learning methods to show its ability. Besides, it is important to investigate the performance of the proposed method using the pre-processing method consisting of LDA and PCA. These methods as feature extraction methods have an undeniable effect on machine learning methods' performance. Therefore, it is predicted that the performance of VAE is also increased using these methods. Finally, the performance of VAE for the combination of sensors and faults should be investigated to find the different aspects of this method. Therefore, the structure of this chapter is divided into seven main sections.

First, the indexes for evaluating the machine learning methods consisting of the confusion matrix and receiver operating characteristics are described. However, most of the analysis in this chapter is based on important indexes consisting of accuracy, precision, recall, and F1-score. Then the sensitivity analysis for hyperparameters of VAE is prepared in Section 2. Next, the evaluation of VAE for all faults separately in rated operating conditions (rated speed and load) is prepared in Section 3. Then the analysis is continued in different operating conditions (different loads and speeds) for different signals in Section 4. The benchmarking of VAE in different conditions is continued in Section 5 using the pre-processing methods. The performance of each method is presented with and without using the pre-processing methods consisting of LDA and PCA. It can be noted that the performance of VAE for mixing selected faults is evaluated in this section as well. In Section 6, the complete analysis is prepared for implementing VAE on different combinations of signals. Besides, in this section, a comparison is made to evaluate using different combinations of sensors or using the pre-processing method to enhance the new method's performance. In the last section of this chapter, VAE is compared to selected machine learning methods consisting of KNN, decision tree, and naïve Bayes in different aspects such as training time and important indexes. Besides, the performance of machine learning methods with and without pre-processing methods is evaluated.

5.1 Evaluation Indexes

One of the essential sections of each algorithm's preparation is evaluating the model's performance. Figure 5-1 shows that evaluating the method is one of the most important stages of preparing a method for work on a data set. Evaluating performance in data science and all processes has a prominent effect on selecting the proposed method. There are many indexes for evaluating the method's performance in data science. The confusion matrix, Receiver Operating Characteristic (ROC), and important indexes consisting of accuracy, precision, recall, and F1-score are the most important indexes for evaluating methods. In the following, they are described, and the important advantages are revealed.

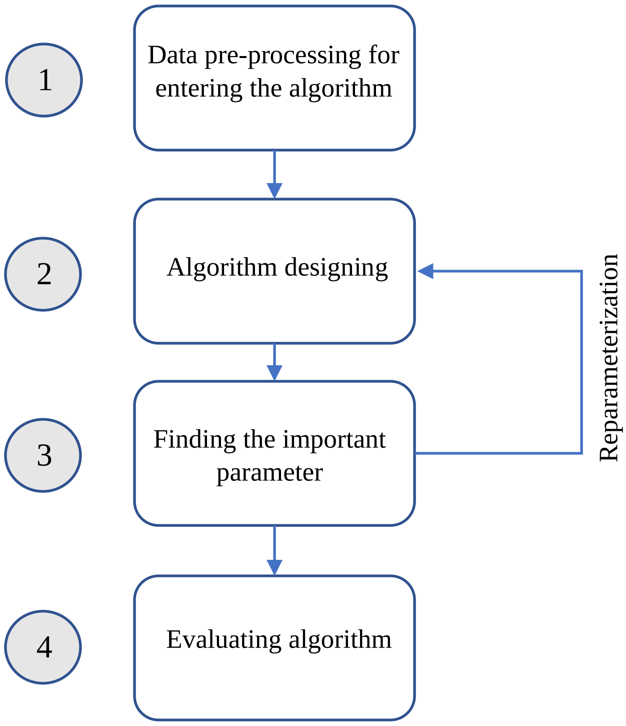


Figure 5-1: The process to evaluate a new method for fault detection

5.1.1 Important Indexes

First, it is needed to introduce some important definitions. As shown in Figure 5-2, there are two series of data dedicated to two different classes; blue and purple data. According to predict each class truly and falsely, there are different definitions as follows:

- True Positive (TP): The positive classes are predicted truly.
- False Negative (FN): The negative classes are misclassified.

- False Positive (FP): The positive classes are misclassified.
- True Negative (TN): The negative classes are predicted truly.

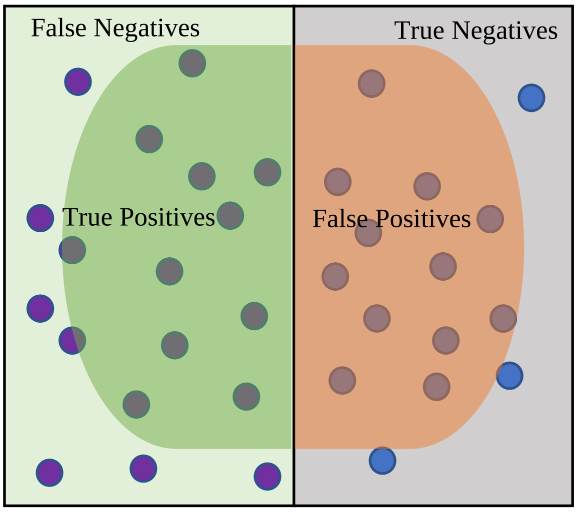


Figure 5-2: A scheme for showing important definitions

These definitions are used for defining different indexes for evaluating machine learning algorithms. The accuracy index is a response to the question of what percent of the positive and negative predictions are correct?" It shows complete knowledge about the machine learning algorithm performance for all classes. This index is defined as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} = \frac{\text{True}}{\text{all of the positive and Negetive}} \tag{5-1}$$

If the dataset is not balanced in different classes, another index called precision can show better method's performance. This index answers this question: "What percent of the positive's predictions are correct?" The following equation shows the calculation of this index:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} = \frac{\text{TP}}{\text{All of the Positive predicted}} \tag{5-2}$$

Precision can get important knowledge about the distribution and data accuracy, and generally, it is better than accuracy for evaluating the performance of an algorithm. Another important metric will be introduced called recall. Recall is also known as sensitivity or true positive rate and is defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} = \frac{\text{TP}}{\text{All of the positive label}} \tag{5-3}$$

Generally, this index is an answer to this question: "What percent of the positive cases did you catch?" Ideally, the recall should be 1 (high) on an excellent classifier. Recall is 1 only when the numerator and denominator are the same. It happens when TP is equal to TP + FN, which means FN is zero. As FN increases, the denominator value becomes larger than the numerator, and the recovery value decreases. Therefore, both precision and recall are needed to be 1 in a good classifier. It means that FP and FN are also zero. Therefore, metrics that consider both accuracy and recall are needed. F1-score is a score that considers both precision and recall and is defined as follows:

$$F_1\text{score} = 2 * \frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (5\text{-}4)$$

Generally, this index responds to this question "What percent of positive predictions were correct?". The F1 score is 1 only if precision and recall are 1. The F1-score is only high when both precision and recall are high. The F1-score is the harmonic mean of precision and recall and is a better measure than precision.

5.1.2 Confusion Matrix

In artificial intelligence, the confusion matrix is a matrix representing the performance of each algorithm. These representations are commonly used in supervised learning algorithms but also unsupervised learning. This matrix in an unsupervised learning algorithm is known as a conformity matrix. However, confusion matrix is used for both of them in the continuation of this research work. Figure 5-3 shows that each matrix column shows a sample of predicted values, and each row contains an actual (correct) instance. The name of this matrix derives from the fact that errors and interferences between results are more easily seen. Outside of artificial intelligence, this matrix is commonly referred to as the contingency matrix or the error matrix. The crucial components of this matrix are True Positive Rate (TPR), False Negative Rate (FNR), False Positive Rate (FPR), and True Negative Rate (TNR), which are calculated as follows:

$$\text{Recall} = \text{Sensetivity} = \text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{True Positive}}{\text{False Positve}+\text{True Positive}} \quad (5\text{-}5)$$

$$\text{FNR} = \frac{\text{FN}}{\text{N}} = \frac{\text{False Negative}}{\text{False Negative}+\text{True Negative}} \quad (5\text{-}6)$$

$$\text{FPR} = \frac{\text{FP}}{\text{P}} = \frac{\text{False Positive}}{\text{False Positve}+\text{True Positive}} \quad (5\text{-}7)$$

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{N}} = \frac{\text{True Negative}}{\text{False Negative+True Negative}} \tag{5-8}$$

This matrix consists of diagonal and non-diagonal components. The diagonal component has a vital role in showing the ability of a method. In other words, the closer this value is to 100%, the algorithm's performance is entirely acceptable.

		Predicted Classes	
		Classes 1	Classes 2
Actual Classes	Classes 1	True Positive Rate (TPR)	False Negative Rate (FNR)
	Classes 2	False Positive Rate (FPR)	True Negative Rate (TNR)

Figure 5-3: Confusion Matrix

5.1.3 Receiver Operating Characteristic

Radar signal analysis during World War II can be considered the first appearance of the Receiver Operating Characteristic (ROC) curve and its applications. Of course, these curves were later used in signal detection theory. After the Pearl Harbor War of 1941, when Americans suffered heavy losses, they decided to improve their radar signals to detect and locate Japanese aircraft. To do this, they measured the ability of a radar receiver to detect an aircraft, and the term "receiver performance" has since been used to evaluate the performance of a signal detection device.

In the 1950s, the ROC curve was also used in psychology to measure human (and sometimes animal) cognitive disorders. In medicine, ROC analysis is widely used to evaluate the accuracy of medical diagnostic tests and determine the tests' accuracy. The ROC curve is also widely used in epidemiologic and medical research. ROC analysis is a common method for evaluating new radiation techniques in radiology. Also, in the social sciences, analysis of the ROC curve is often referred to as the "ROC accuracy ratio" and is a standard method of evaluating predictive probabilistic models. The ROC curve and its applications have proven helpful for machine learning, especially in the supervised field. Therefore, the values of the ROC curve can be the basis for comparing and evaluating classifier algorithms. ROC curves are also used to validate predictions from meteorology. Generally, ROC is a graph to show the ability to evaluate a classification system with a variable detection threshold. The ROC curve is created by plotting the ratio or "True Positive Rate", abbreviated TPR, to the "False Positive Rate" (FPR). However, the threshold for these values is variable. Because of this, a continuous graph is created. In general, provided the probability distribution for both (TPR) and (FPR) parts is known, the ROC curve is obtained if the "cumulative distribution function" or the area below the distribution curve determines the probability of correctly detecting the signal in considering the vertical axis and the cumulative distribution function of incorrect signal recognition in the horizontal axis.

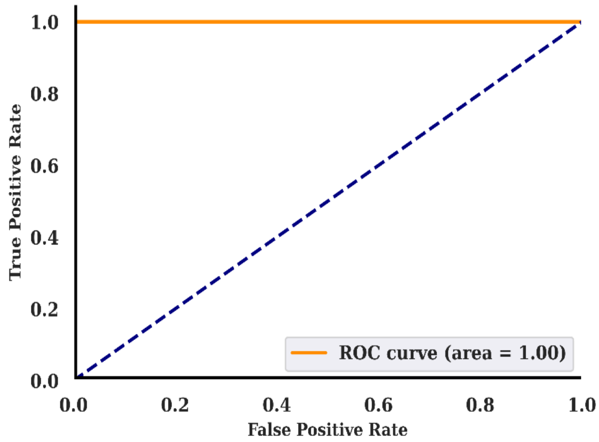


Figure 5-4: The ROC curve

The ROC curve is also known as the "Relative Operating Characteristic" because it compares two modes of operation (TPR, and FPR), as shown in Figure 5-4. Thus, the curve is known nowadays (regardless of the concept of signal recognition by the receiver) as the ROC curve is a graph of relative performance characteristics. This procedure can give a proper perspective on different aspects of each method. According to the investigation of different methods for evaluating the machine learning method, it can be understood that the confusion matrix gives the best view of the classification situation of a method, and it can be used for analysis in the following. However, when the number of classes rises, it is better to use essential indexes instead of a confusion matrix to get a general overview of the performance of VAE for all methods. Besides, using ROC prolongs this thesis because it is needed to draw the ROC curve for each class in each analysis. In fact, ROC is the proper method for two classes of data. In the following, the analysis of the proposed method in different aspects is presented. First, the sensitivity analysis of VAE’s hyperparameters is investigated. Then the performance of VAE for detecting the different severity of faults in rated speed and load is evaluated. Besides, the performance of VAE for 32 classes in different loads and speeds is presented in the next step. In the next section, the impact of using pre-processing method consisting of LDA and PCA is investigated. Besides, the performance of VAE is evaluated for selected mixing faults. In the next section, the performance of VAE using the different combinations of the sensor is evaluated. Finally, some of the machine learning methods are compared with VAE for different datasets from different aspects such as training time and important indexes.

5.2 Variational Autoencoder's Hyperparameters

In this section, the influence of selecting each parameter on the performance of VAE is investigated thoroughly. As was described, it is needed to tune different hyperparameters in the structure of VAE. It can be helpful to note that analyzing this section for investigating the sensitivity of different sensors is just for showing the importance of each parameter on network performance. It is impossible to find each optimized parameter value using this analysis. The best way to find the proper value of parameters for each signal is to investigate different scenarios for all of the parameters. However, this work needs much memory capacity for the system and is more time-consuming. Therefore, the network's sensitivity to each hyperparameter is investigated in this research work.

The important hyperparameters of VAE consist of input size, hidden dimension size, batch size, epochs, and the number of neurons per layer. Each deep learning network is constructed from many layers. According to different situations, it is needed to use different layers with different neurons. An important point to note is that increasing the number of layers and neurons causes over-training of the network. Therefore, it is necessary to have a trade-off between the proper number of layers, neurons, and to prevent over training in the network.

In the continuation of this research work, the VAE with three layers for the encoder and three layers for the decoder is selected. This number of layers are found according to experience of researcher. Besides, when the number of layers increased the chance of overtraining is increased. Therefore, it is found that increasing more than three layers increases the chance of overtraining.

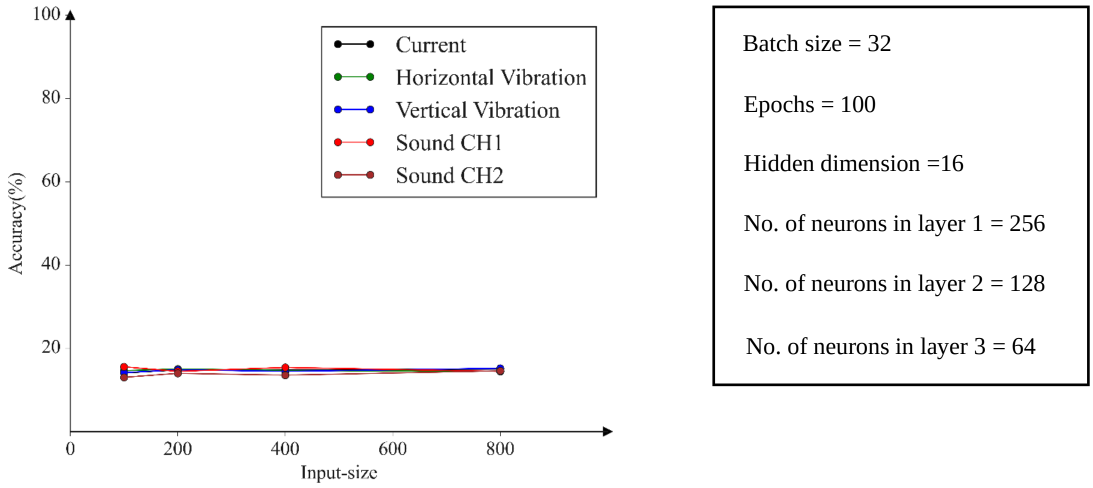


Figure 5-5: The relation between the input size and accuracy

The first parameter for investigating is the input size. The input size is the number of data prepared to enter the network. As in deep learning, the data dimension is high, and it is impossible to analyse all of the data once. Thus, it is better to prepare the data in a different category for the network. As shown in Figure 5-5, the network performance for increasing the input size of data for different sensors is evaluated

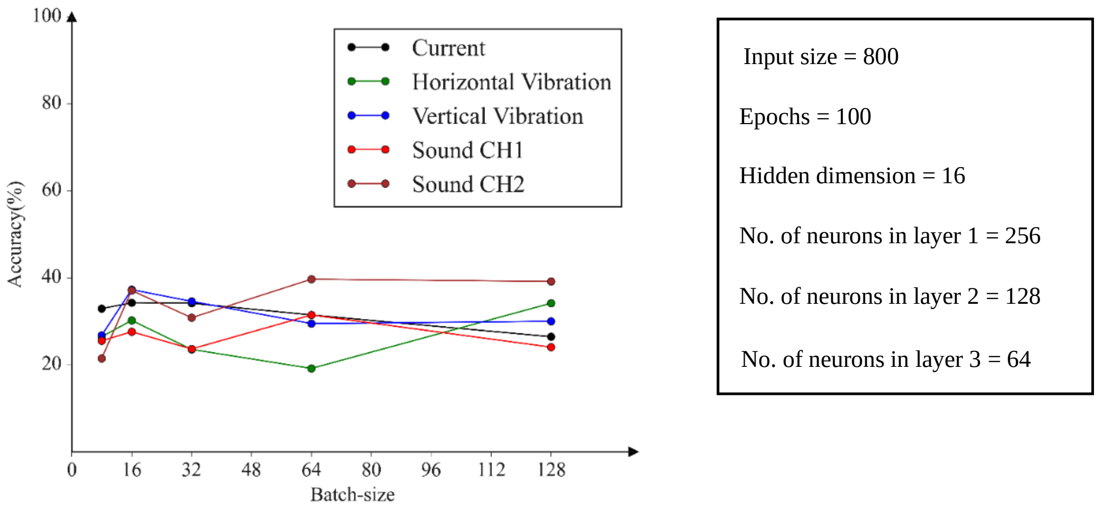


Figure 5-6: The relation between the batch size and accuracy

. It is found that increasing input size does not affect the network's performance for each sensor. However, when the parameters of the network change, there is a possibility that the effect of input size is increased. It is important to know that input size should not be mixed with batch size. Batch size is defined to show how many data is needed for the training network by each iteration.As shown in Figure 5-6, the network's performance for the different batch sizes for each sensor is different. As shown, increasing the batch size does not directly relate to the performance of VAE. Therefore, it is not possible to get a complete conclusion according to this result for different signals. The parameter that has an essential effect on the performance of the network and training time is the number of iterations for each training, called the epochs number. As shown in Figure 5-7, when the number of epochs increases, the network's performance will be enhanced approximately. However, this idea is not valid for sound signals in channel one. The performance of a VAE for sound signal in channel one with 150 epochs is better than 200 epochs. Besides, the epochs number has a vital role in over-training the network. It is better to have a trade-off between increasing the number of epochs and preventing over-training.

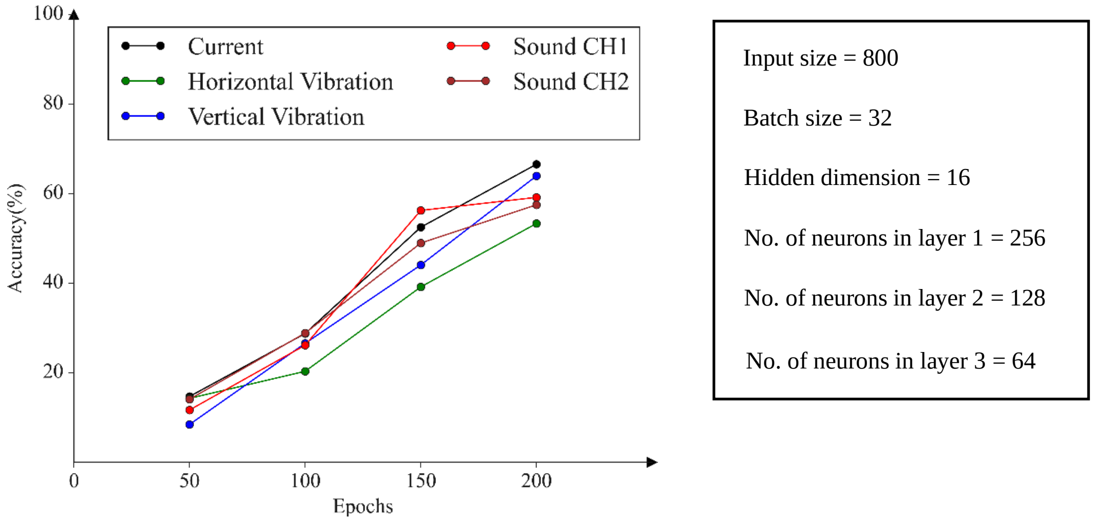
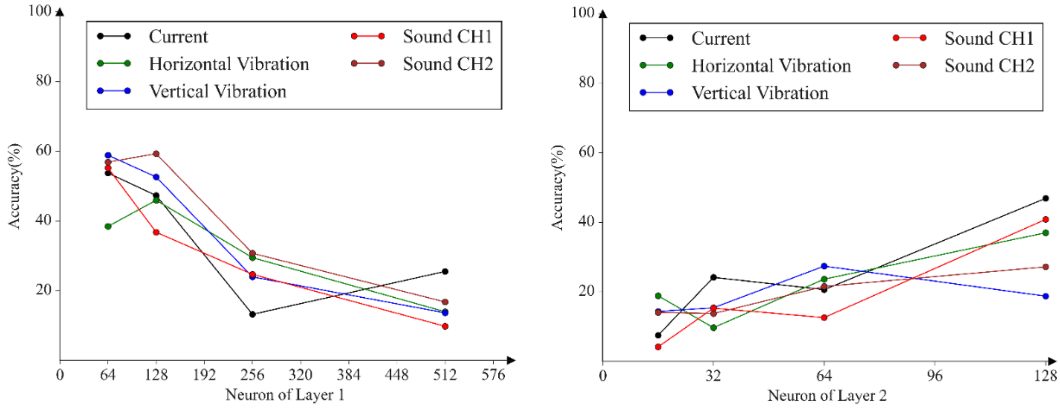
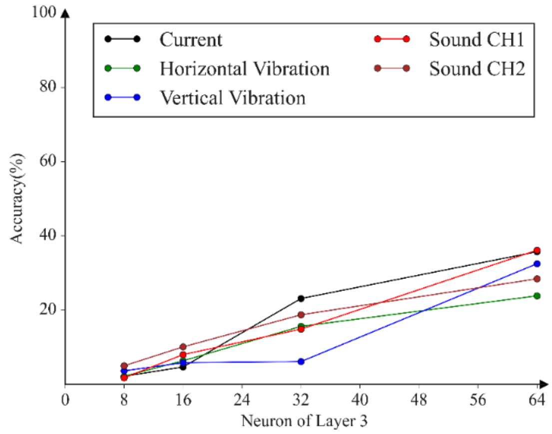


Figure 5-7: The relation between the epochs and accuracy

Each layer has a different number of neurons; these neurons receive the input, and after applying the activation function, the input of the next layer is produced. According to the structure of VAE, first, the data using an encoder reduces dimensionality, and the number of neurons per layer should have a decreasing trend. For instance, if the neurons in layer one is 256, using 512 neurons in the next layer is not possible. It is better to use a lower number of neurons compared to previous layers. However, the number of neurons in each layer for the decoder section of VAE should have an increasing trend. Therefore, the analysis of VAE for different layers are done according to this number of neurons for different layers. For instance, when the changing of neurons in layer one is investigated, the number of neurons in other layers is 128, and 64 for layers number two and three, respectively. Figure 5-8 illustrates the effect of different number of neurons on VAE’s accuracy for different layers. As shown in Figure 5-8 (a), increasing the number of neurons in layer one causes VAE's performance to be reduced. However, this analysis is not true for the current signal entirely. As can be seen, the performance of VAE for the current signal with 256 neurons in layer one is better than 512 neurons. The performance of VAE based on changing the number of neurons in layer two is shown in Figure 5-8 (b). It is clear that increasing the number of neurons has a proper effect on VAE's performance. The performance of neurons number in layer three is shown in Figure 5-8 (c), which increasing the neurons' number causes to enhance the VAE's performance. It can be noted that the increasing and decreasing trend of the number of neurons is based on the restriction in the encoder and decoder section.



(a) (b)



Input size = 800

Batch size = 32

Epochs = 100

Hidden dimension =16

No. of neurons in layer 1 = 256

No. of neurons in layer 2 = 128

No. of neurons in layer 3 = 64

(c)

Figure 5-8: The relation between the neuron of layer and accuracy
(a) Layer 1 (b) Layer 2 (c) Layer 3

After successive layers of the encoder, there is a hidden layer with different neurons whose task is to receive the data from the encoder, produce the new data according to the defined distribution, and send it to the decoder part to reconstruct the data. Like other layers, this layer has some neurons; this layer's neurons called hidden dimensions, are essential for VAE. Figure 5-9 shows the performance of VAE for increasing the hidden dimension. As it can be seen, the 16 neurons are the proper amount for horizontal, vertical vibration, and sound signal for channel two. However, 12 is the proper amount for the current and sound channel two. Therefore, different amounts of hidden dimensions are proper for different signals, and it is not possible to find a specific trend for VAE.

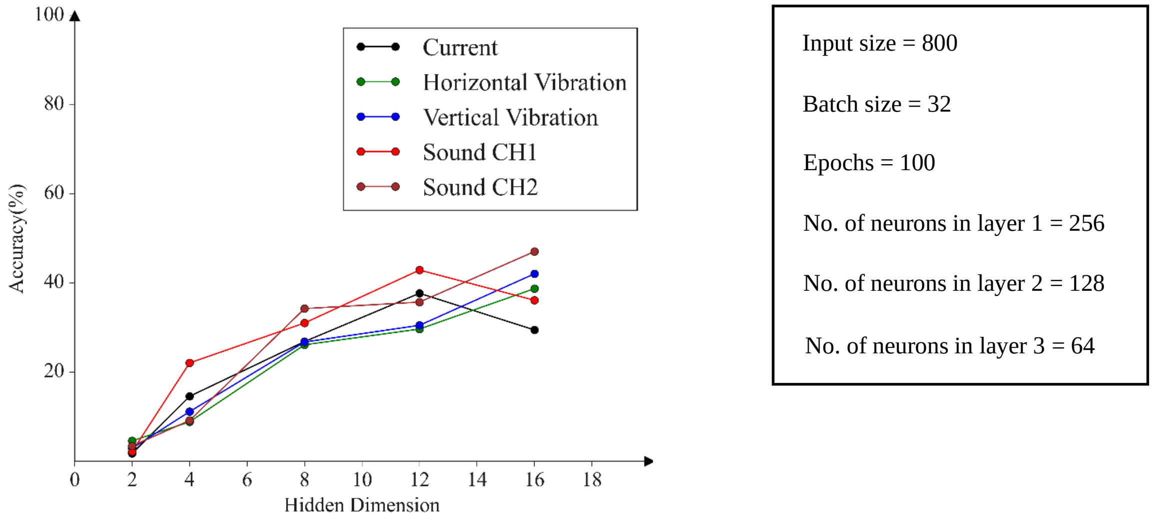


Figure 5-9: The relation between the hidden dimension and accuracy

In conclusion, finding the optimised values of each hyperparameter based on investigating its effects on the performance of the VAE is not an easy work. There are several techniques that can be used to find proper hyperparameters for neural networks. Grid search involves evaluating all possible combinations of hyperparameters within defined ranges, while random search randomly samples from those ranges. Bayesian optimization uses probabilistic models to focus on the most promising hyperparameters, while gradient-based optimization applies gradient descent to hyperparameters. The best approach depends on the specific problem and resources available, and it's often a good idea to try multiple approaches and compare their results. Because of memory constraints, it is not feasible to simultaneously explore all values for various hyperparameters using any of these techniques. However, it is possible to find the optimised value of each parameter approximately based on the experience of the neural network designer. In this research work, the structure of VAE is based on 3 layers for the encoder and 3 layers for the decoder, and the parameters of VAE are presented in Table 5-1.

Table 5-1: The parameters of VAE

Parameter	Value
Input size	800
Batch size	32
Epochs	100
No. of neurons in layer 1	256
No. of neurons in layer 2	128
No. of neurons in layer 3	64
Hidden dimension	16
Sample rate	16 kHz
Time	5 s

One of the powerful Graphical Processor Unit (GPU) models, TESLA V100-8Q, manufactured by NVIDIA, was used to analyze the data in this research work. This type of processor uses Compute Unified Device Architecture (CUDA) in its structure for processing. This GPU known as one of the high-ranked processors is evaluated with a 7.0 computational capability' mark.

5.3 Variational Autoencoder Analysis in Rated Condition

The VAE is evaluated for classifying different faults separately in rated speed and rated load. One of the critical abilities of proper diagnosis methods is severity estimation. Therefore, the ability of VAE to separate the different severity of different faults is investigated in the first step of this section. Then, its performance for different severity is evaluated in other loads and speeds condition.

5.3.1 Analysing VAE for Demagnetization Fault

This section presents the classification of demagnetization fault with various severity using VAE in rated condition. In the first step, it is important to evaluate the overtraining situation. When the loss and validation loss start to diverge by increasing the iteration, the overtraining happens. It means the network performs properly in classifying the training dataset though its performance in detecting new data in the test dataset is improper. Therefore, the network's loss and validation loss are investigated to know the overtraining situation in the network. As shown in Figure 5-10 (a), the loss and validation loss are decreased by increasing the iteration. Besides, these two curves are not diverged, and any symptom of overtraining cannot be seen. Also, when the iteration increases, the accuracy of the network with many swings progresses and finally gets a proper value, as shown in Figure 5-10 (b). Therefore, it can be found that the network has a proper performance and does not enter the overtraining situation.

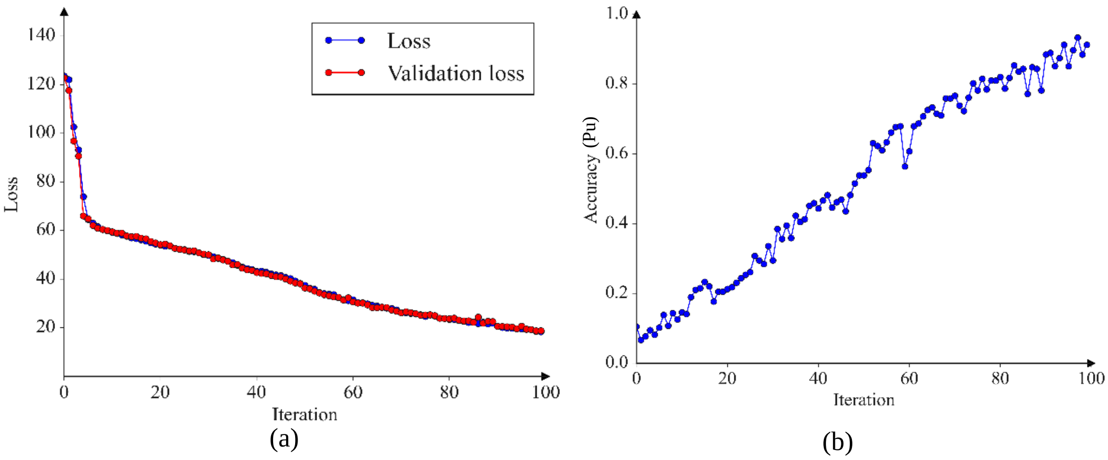


Figure 5-10: Evaluating of VAE for the current sensor using loss and accuracy for different iteration (a) Loss and validation loss (b) Accuracy

The evaluation indexes for different classes are separately presented in Table 5-2. It performs properly with 92 %, 93%, and 93% on average for precision, recall, and F1-score, respectively. However, the network in detection faults No. 4 and 8 has no proper performance. Fault No. 4 has 78% precision, which means that the network for classifying the true positive sample among the positive predicted classes has an improper performance; however, the VAE for detecting the true positive samples among the positive classes is 51%. These results show that the sensitivity of VAE in class No. 4 is improper. Also, this class's FNR, FPR, and specificity or TNR are 49%, 49%, and 51%, respectively.

The confusion matrix for the current signal is presented in the following. As shown in Figure 5-11, the performance of each class can be seen based on the diagonal members and non-diagonal members of the matrix. The confusion matrix results can approve the results of Table 5-1. As seen, the results of class No. 4 are not proper, and the result of this class with 51% precision is distinguished by the light blue colour from the other diagonal members. Also, other essential indexes can be extracted from the confusion matrix in Figure 5-11. For instance, the sensitivity or recall according to equation 5-3 for class No. 4 is calculated using the confusion matrix in Figure 5-11. For calculating the recall, the number of diagonal members of class No. 4 is divided by the sum of the amount in the column of class No. 4, and finally, 78 % is achieved. It can be helpful to present formulas to calculate precision and recall using a confusion matrix as follows:

$$Precision_i = \frac{M_{ii}}{\sum_j M_{ji}} \tag{5-9}$$

$$Recall_i = \frac{M_{ii}}{\sum_j M_{ij}} \tag{5-10}$$

Where M_{ii} is dedicated to diagonal members of the confusion matrix and M_{ij} is for the non-diagonal member in the confusion matrix. In the continuation of this research work, the confusion matrix is the primary tool for evaluating networks' ability. Since using the confusion matrix makes it possible to investigate the network's ability for each class and calculate the evaluation indexes, the same information in different tables and figures are avoided. Besides, this approach causes to avoid prolonging the analysis for each signal.

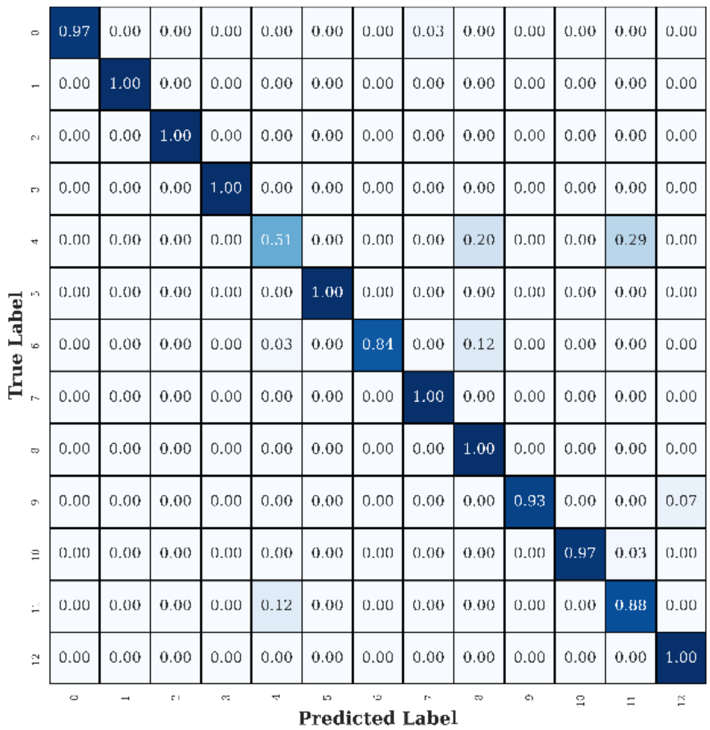
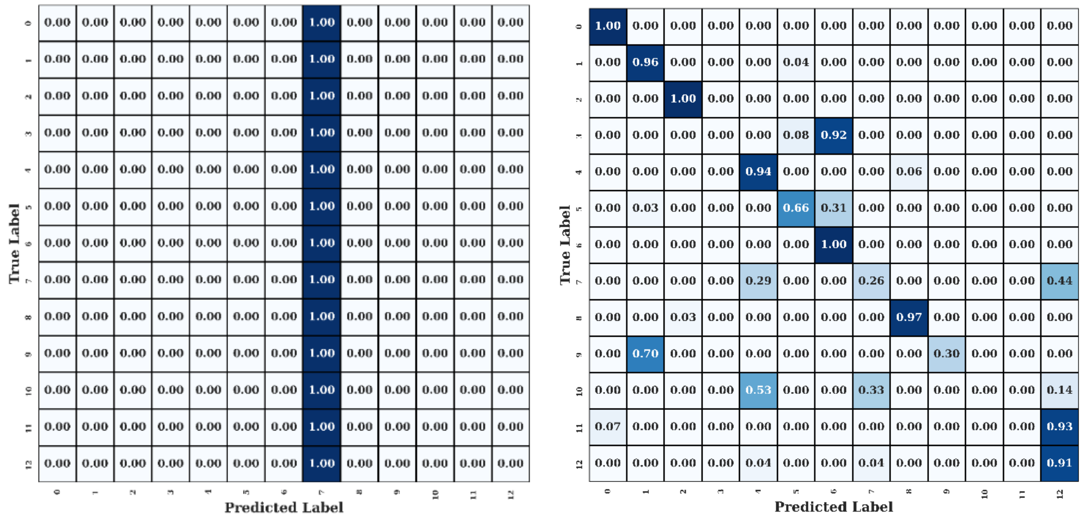


Figure 5-11: Confusion matrix for detecting demagnetization using current sensor in per unit

In the following, the results for vibration and sound signals are presented. Figure 5-12 presents the confusion matrix for both vibration direction signals. As shown in Figure 5-12 (a), the performance of VAE for the horizontal sensor is not proper, the probability distribution of data in this sensor is not well, and the network for finding all of the faults has some problems. The accuracy of this sensor is 6%, and the average of precision, recall, and F1-score based on equations (5-2), (5-3), and (5-4) are 0%, 8%, and 1%, respectively. However, the performance of vertical vibration shown in Figure 5-12 (b) is better than the horizontal sensor. The accuracy of this sensor is 58% and the average precision, recall, and F1-score indexes are 52%, 62%, and 52%, respectively. According to the reality that electrical machine in different direction has vibration with different amplitudes, and probably the effect of fault in one direction is more powerful than in others directions, VAE using vibration signal is not so successful in one direction.

Table 5-2: The results of important indexes for current sensor demagnetization fault in different orientation and severity in per unit

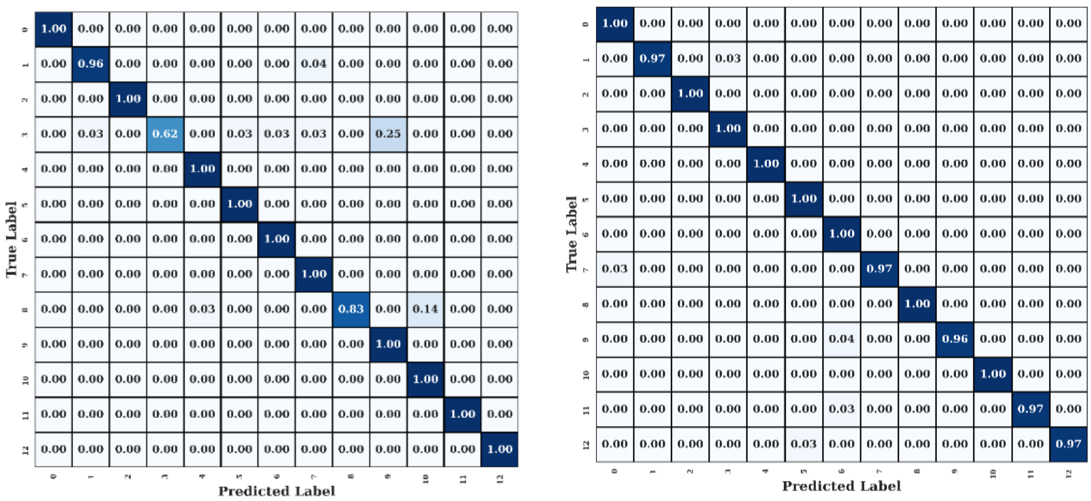
Classes		Severity and orientation of demagnetization	Precision	Recall	F1-Score
Name	Number				
Normal	0	-	1	0.97	0.98
Full-50205 steel	1	100%	1	1	1
Full-No Magnet	2	100%	1	1	1
x-50201 steel	3	x-20%	1	1	1
x-50202 steel	4	x-40%	0.78	0.51	0.62
x-50203 steel	5	x-60%	1	1	1
y-501005 steel	6	y-50%	1	0.84	0.92
y-501505 steel	7	y-75%	0.97	1	0.98
y-500505 steel	8	y-25%	0.65	1	0.78
z-10205 steel	9	z-20%	1	0.93	0.96
z-20205 steel	10	z-40%	1	0.97	0.98
z-30205 steel	11	z-60%	0.73	0.88	0.80
z-40205 steel	12	z-80%	0.94	1	0.97
Average			0.92	0.93	0.93



(a) (b)

Figure 5-12: Confusion matrix for detecting demagnetization using vibration sensor in per unit (a) Horizontal vibration (b) Vertical vibration

As shown in Figure 5-13 (a), the performance of VAE for the sound signal in channel one is proper with 95% accuracy. Also, the performance of sound signal in channel two, shown in Figure 5-13 (b), is better than sound channel one. The accuracy of this sensor is 99%, and the average precision, recall, and F1-score according to equations (5-2), (5-3), and (5-4) are 99%, 99%, and 99%, respectively. Therefore, the performance of VAE using sound signals, compared to current and vibration sensors used in this research work, have a better ability to diagnose demagnetization in rated condition.



(a) (b)

Figure 5-13: Confusion matrix for detecting demagnetization using a sound sensor in per unit (a) Sound sensor CH1 (b) Sound sensor CH2

Table 5-3: The results of essential indexes for current sensor and bearing faults in the different orientation and severity in per unit

Classes		Depth and width of penetration	Precision	Recall	F1-Score
Name	Number				
Ball-1	0	0.1 mm-1 mm	0.03	0.03	0.03
Ball-1.5	1	0.1 mm-1.5 mm	0.72	0.72	0.72
Ball-1.25	2	0.1 mm-1.25 mm	0	0	0
Ball-1.75	3	0.1 mm-1.75 mm	0.36	0.75	0.48
Ball-2	4	0.15 mm-2 mm	1	0.42	0.59
Inner-1	5	0.1 mm-1 mm	0.6	0.48	0.53
Inner-1.5	6	0.1 mm-1.5 mm	0.18	0.48	0.26
Inner-1.25	7	0.1 mm-1.25 mm	0.70	0.78	0.74
Inner-1.75	8	0.1 mm-1.75 mm	0.78	0.78	0.78
Inner-2	9	0.15 mm-2 mm	0	0	0
Normal	10	-	0.38	0.17	0.23
Outer-1	11	0.1 mm-1 mm	1	1	1
Outer-1.5	12	0.1 mm-1.5 mm	0.54	0.70	0.61
Outer-1.25	13	0.1 mm-1.25 mm	0.76	0.93	0.84
Outer-1.75	14	0.1 mm-1.75 mm	0.41	0.70	0.52
Outter-2	15	0.15 mm-2 mm	1	1	1
Average			0.53	0.56	0.52

5.3.2 Analyzing VAE for Bearing Fault

Bearing faults are more likely to occur in comparison to demagnetization and eccentricity. Also, this fault can occur in all electrical machines compared to the demagnetization fault. According to Chapter 4, 15 types of bearing faults are considered on different bearing components such as inner ring, outer ring, and ball bearing for simulating different severity of the fault, as shown in Table 5-3. In the first step, same as the previous section, the network for overtraining is investigated. As shown in Figure 5-14 (a), the loss and validation loss decrease by increasing the iteration, and the validation loss takes a route the same as loss and is not diverged from the loss validation curve. Therefore, the network is not located in an overtraining situation. Furthermore, as shown in Figure 5-14 (b), the accuracy after iteration No. 40 is around 50%, and this is not a proper situation.

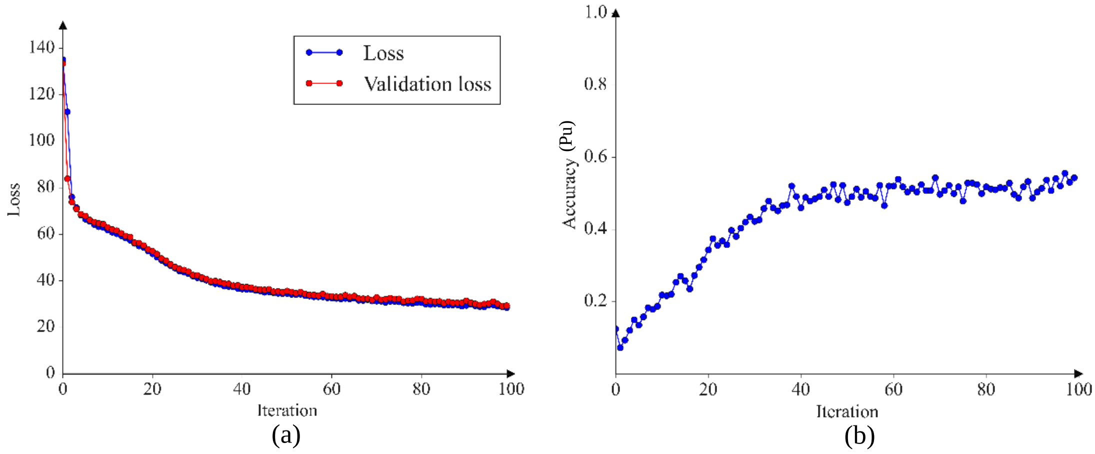


Figure 5-14: Evaluating of VAE for the current sensor using loss and accuracy for different iterations for bearing faults (a) Loss and validation loss (b) Accuracy

Table 5-3 presents the different classes of bearing faults with different severity. According to Figure 5-15, the confusion matrix for the current sensor is presented, and it shows that the performance of the current signal for detecting different severity of faults is not good with 54% accuracy. Therefore, It is better to use the pre-processing method on this data for better results. The vibration and sound signals are investigated in the following.

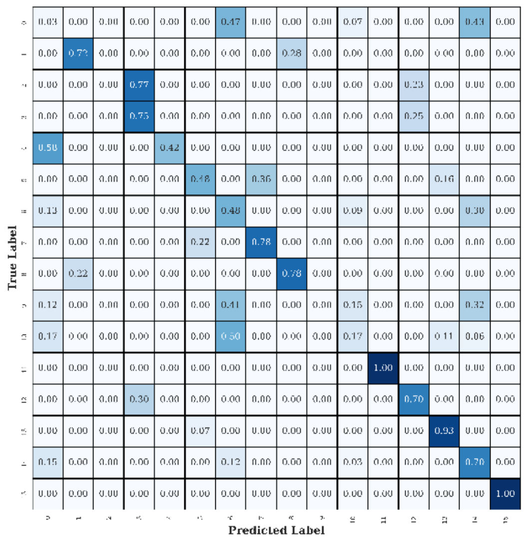
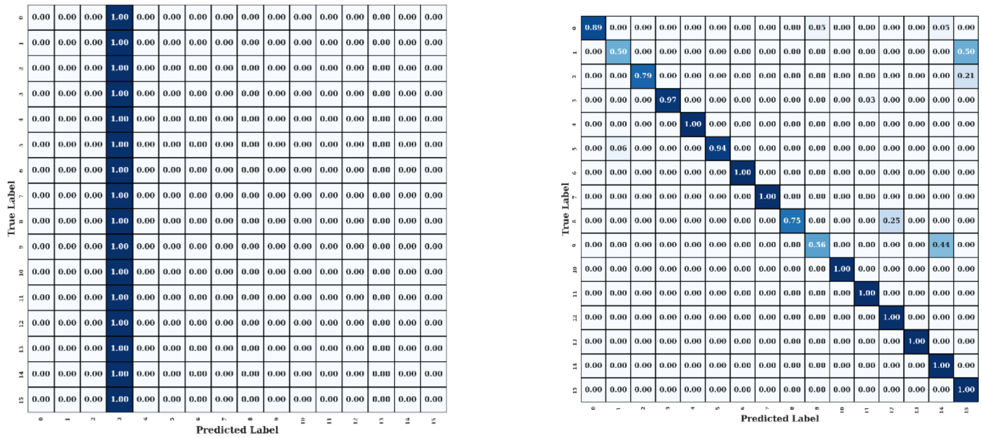


Figure 5-15: Confusion matrix for detecting bearing faults using current sensor in per unit

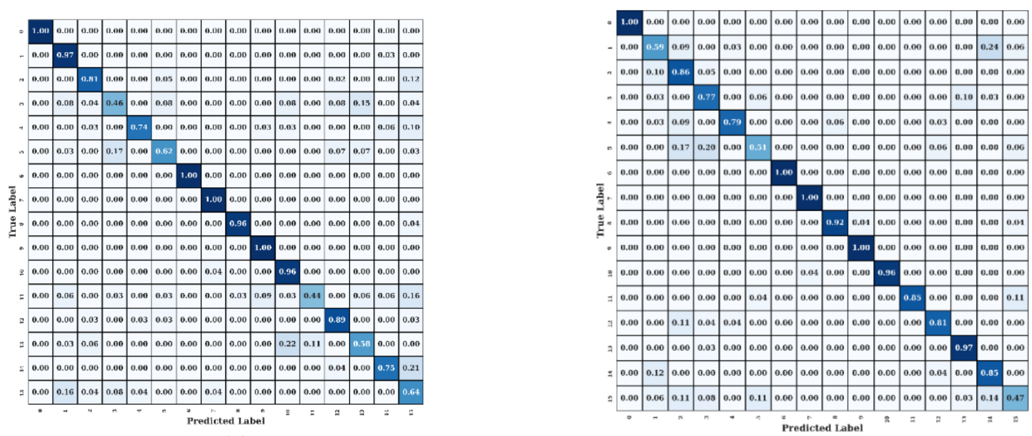
As shown in Figure 5-16 (a), the performance of VAE for the horizontal vibration sensor is not proper, the probability distribution of data in this sensor is not well, and the network for detecting all of the faults has not a proper situation. However, Figure 5-16 (b) shows that the vertical vibration performance is better than the horizontal vibration. The significant result from the analysis of vibration signal is that vibration sensor in horizontal orientation is not proper for diagnosing the bearing fault because of the higher effect of fault in this direction compared to other directions.



(a) (b)

Figure 5-16: Confusion matrix for detecting bearing faults using vibration sensor in per unit (a) Horizontal vibration (b) Vertical vibration

This diagnosis situation is the same for demagnetization fault using a horizontal vibration sensor. Figure 5-17 presents the confusion matrix for both sound channels. As shown in Figure 5-17 (a), the performance of VAE for sound channel one is proper. Also, the performance of sound in channel two is better than sound channel one as shown in Figure 5-17 (b) with 82% accuracy. The critical result from the analysis of bearing faults is that vertical vibration, sound channel one, and two signals, compared to the current sensor data used in this research work, have a better ability to diagnose a bearing fault in rated speed and rated load.



(a) (b)

Figure 5-17: Confusion matrix for detecting bearing faults using a sound sensor in per unit (a) Sound sensor CH1 (b) Sound sensor CH2

5.3.3 Analyzing VAE for Eccentricity Fault

In this section, the results of evaluating VAE for eccentricity, one of the most important faults that causes the non-uniformity in the air gap is presented. This fault directly relates to one of the crucial faults occurring more frequently in high-power machines called the unbalanced rotor fault. According to Chapter four, various severity of static eccentricity are considered as shown in Table 5-4 for analyzing.

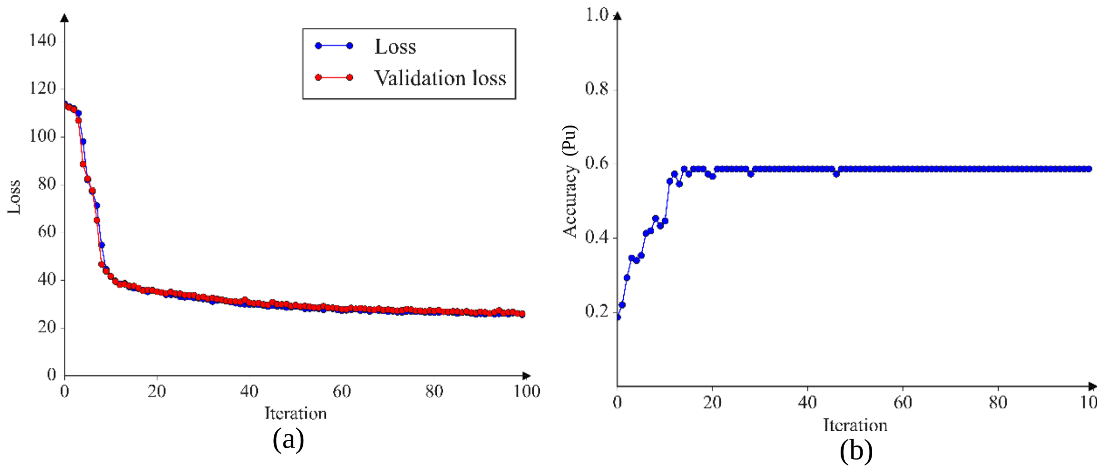


Figure 5-18: Evaluating of VAE for the current sensor using loss and accuracy for different iterations for eccentricity faults (a) Loss and validation loss (b) Accuracy

In the first step, same as the previous sections, the network is checked for overtraining. As shown in Figure 5-18 (a), the loss and validation loss decrease by increasing the iteration, and the validation loss takes a route the same as loss and is not separated from the loss curve. Therefore, the network is not located in an overtraining situation. Besides, Figure 5-18 (b) shows that the curve's accuracy after iteration No. 20 is located around 0.6. Also, Table 5-4 shows the results of important indexes for different classes' eccentricity faults for different fault severities. According to Figure 5-19, the confusion matrix for the current sensor is presented, which shows that the performance of the current signal for finding different severity of faults is not proper. The accuracy of this sensor is 59%, and the average precision, recall, and F1-score according to equations (5-2), (5-3), and (5-4) are 69%, 58%, and 60%, respectively. As shown in Figure 5-19 and Table 5-4, VAE in class No. 0 is located in a not proper situation and cannot detect this fault well. Therefore, the VAE using the current sensor has not a proper performance for diagnosing the bearing faults with different severity using VAE in this research work.

Table 5-4: The results of important indexes for current sensor eccentricity fault in different severity in per unit

Classes		Severity	Precision	Recall	F1-Score
Name	Number				
Ecc-0.1	0	0.1 mm	0.29	0.48	0.37
Ecc-0.15	1	0.15 mm	0.71	0.73	0.72
Ecc-0.2	2	0.2 mm	1	0.71	0.83
Ecc-0.25	3	0.25 mm	0.45	0.48	0.46
Normal	4	-	1	0.48	0.65
Average			0.69	0.58	0.60

The vibration and sound signal are investigated in the following. As shown in Figure 5-20 (a), the performance of VAE for the horizontal vibration sensor is proper. However, Figure 5-20 (b) shows the vertical vibration performance, which is not proper compared to the horizontal vibration sensor because of the higher effect of fault in one direction compared to other directions. The accuracy of this sensor is 72%, and the average precision, recall, and F1-score according to equations (5-2), (5-3), and (5-4) are 81%, 76%, and 71%, respectively. In contrast to demagnetization and bearing faults, horizontal vibration signal for diagnosing eccentricity faults have better performance.

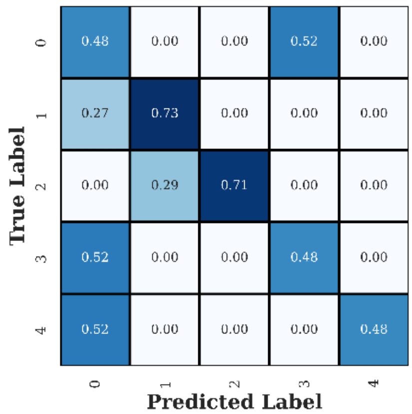
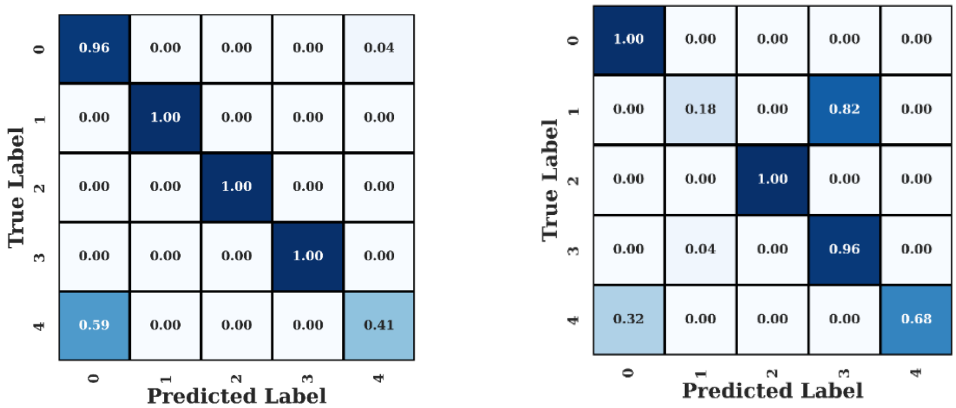


Figure 5-19: Confusion matrix for detecting eccentricity faults using current sensor in per unit

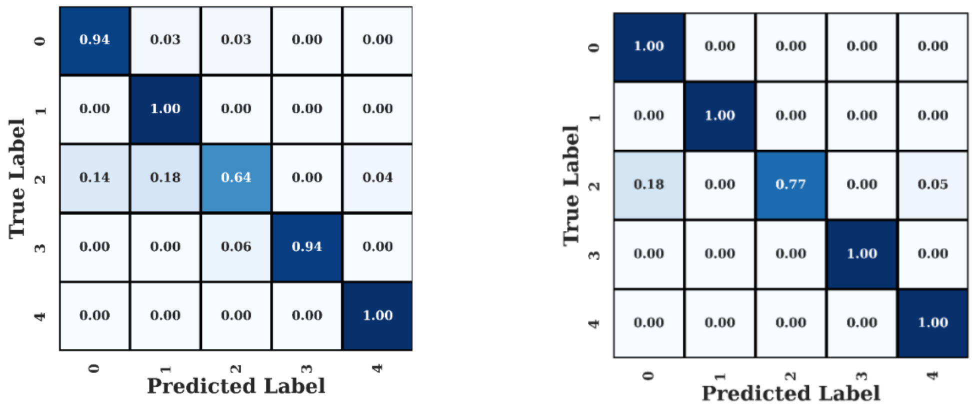
Therefore, the horizontal and vertical vibration sensors are proper for detecting eccentricity faults with different severity. Figure 5-21 presents the confusion matrix for both sound signals. As shown in Figure 5-21 (a), the performance of VAE for sound signal in channel one is proper. Also, the performance of the sound signal in channel two, shown in Figure 5-21 (b), is better than sound in channel one. The accuracy of this sensor is 97%, and the average precision, recall, and F1-score according to equations (5-2), (5-3), and (5-4) are 97%, 95%, and 96%, respectively. It can be found that sound signal in channel two has a proper performance compared to other sensors for detecting eccentricity faults in rated condition. Sound signals among all signals for various faults have better accuracy without pre-processing.



(a) (b)

Figure 5-20: Confusion matrix for detecting eccentricity faults using vibration sensor in per unit (a) Horizontal vibration (b)Vertical vibration

To conclude, the performance of VAE for different faults using different sensors is evaluated. The performance of the VAE using sound signals compared to all of the signals for all faults is the best. Besides, the current sensor's performance is proper for the demagnetization fault and the vibration signal for the eccentricity fault. As demagnetization has a low effect in the direction in which the vibration signals are gathered, VAE without any pre-processing has not a proper performance.



(a) (b)

Figure 5-21: Confusion matrix for detecting eccentricity faults using a sound sensor in per unit (a) Sound sensor in CH1 (b) Sound sensor in CH2

Table 5-5: The evaluation of important indexes for 32 classes using the current signal for rated speed and different loads in per unit

Speed	Load	Accuracy	Precision	Recall	F1-Score
100%	No	0.84	0.87	0.85	0.84
	20%	0.85	0.85	0.85	0.84
	40%	0.84	0.85	0.85	0.84
	60%	0.82	0.83	0.83	0.82
	80%	0.63	0.68	0.65	0.67
	100%	0.68	0.66	0.67	0.65
	120%	0.73	0.71	0.74	0.71

5.4 Variational Autoencoder Analysis in Different Conditions

In pervious sections, the performance of the VAE for each fault and each sensor were presented in rated conditions. This section is dedicated to present the results of implementing VAE in different speeds and loads. In contrast to the previous sections, the performances of the VAE are investigated for all types of faults together with the same VAE. The input data consist of demagnetization (12 classes), bearing (15 classes) and eccentricity (4 classes) faults, and the normal situation (1 class). As Table 5-5 shows the important indexes results for the current signal in rated speed and different loads. According to this table, the algorithm's ability to detect faults is not proper for higher loads than 80%. The average accuracy for the current sensor is 77.14%.

Table 5-6: The evaluation of important indexes for 32 classes using the current signal for rated load and different speeds in per unit

Load	Speed	Accuracy	Precision	Recall	F1-Score
100%	20%	-	-	-	-
	40%	0.85	0.88	0.85	0.85
	60%	0.69	0.68	0.71	0.68
	80%	0.94	0.93	0.94	0.93
	100%	0.68	0.66	0.67	0.65
	120%	0.76	0.79	0.77	0.75

Also, Table 5-6 shows the results of implementing VAE on data for rated load and different speeds for the current signal. It is clear that for 60%, 100%, and 120% speed, the algorithm's performance is not proper. The average accuracy for this analysis for the current sensor is 78.40%. It is not possible to implement 20% speed and 100% load scenario on the motor because the motor cannot produce enough torque for this load. Therefore, the evaluation results for 100% load and 20% speed is not presented in Table

5-6. As shown in Table 5-7, the result for rated speed and different loads for the horizontal vibration signal is presented. The results show that the performance of the VAE for low and high loads is not proper. Therefore, feature extraction is needed for different load ranges. The average accuracy for this situation is 67.14%. As shown in Table 5-8, the performance of VAE for detecting faults for different speeds is better than different loads in Table 5-7. Also, at low speed, the performance of the network is better than at other speeds. Besides, when the speed rises, in some case the situation for diagnosis is proper and for some cases, it is not proper. The average accuracy for this situation is 79.20% as well.

Table 5-7: The evaluation of important indexes for 32 classes using the horizontal vibration signal for rated speed and different loads in per unit

Speed	Load	Accuracy	Precision	Recall	F1-Score
100%	No	0.40	0.50	0.42	0.33
	20%	0.82	0.83	0.82	0.80
	40%	0.84	0.83	0.83	0.82
	60%	0.86	0.86	0.87	0.86
	80%	0.67	0.67	0.67	0.64
	100%	0.61	0.60	0.61	0.59
	120%	0.71	0.73	0.72	0.71

Table 5-8: The evaluation of important indexes for 32 classes using the horizontal vibration signal for rated load and different speed in per unit

Load	Speed	Accuracy	Precision	Recall	F1-Score
100%	20%	-	-	-	-
	40%	0.90	0.93	0.90	0.89
	60%	0.70	0.70	0.71	0.68
	80%	0.95	0.93	0.95	0.95
	100%	0.61	0.60	0.61	0.59
	120%	0.80	0.79	0.80	0.79

To conclude, for horizontal vibration, in different loads at rated speed and different speeds with a rated load, the network's performance is not proper because of the higher effect of fault in one direction compared to other directions, and it is needed to implement feature extraction methods. In the following of this section, important vertical vibration indexes for different loading conditions, and speeds are investigated. Table 5-9 shows the results for rated speed and different loads from no-load to 120%

load. As can be seen, the network's performance for no-load is improper, and it is not possible to use VAE on no-load conditions. Similarly, overload situation is the same for VAE performance. However, the performance of the proposed method in middle loads is appropriate. One important point to note is that the performance of VAE for horizontal and vertical vibration signals in the no-load case is the same.

Table 5-9: The evaluation of important indexes for 32 classes using the vertical vibration signal for rated speed and different loads in per unit

Speed	Load	Accuracy	Precision	Recall	F1-Score
100%	No	0.53	0.55	0.55	0.51
	20%	0.85	0.85	0.85	0.84
	40%	0.85	0.85	0.85	0.83
	60%	0.83	0.85	0.84	0.84
	80%	0.70	0.72	0.70	0.69
	100%	0.71	0.72	0.71	0.68
	120%	0.67	0.69	0.69	0.65

Table 5-10: The evaluation of important indexes for 32 classes using the vertical vibration signal for rated load and different speeds in per unit

Load	Speed	Accuracy	Precision	Recall	F1-Score
100%	20%	-	-	-	-
	40%	0.92	0.92	0.92	0.92
	60%	0.64	0.67	0.66	0.63
	80%	0.88	0.83	0.88	0.84
	100%	0.71	0.72	0.71	0.68
	120%	0.74	0.77	0.76	0.72

In Table 5-10, The performance of vertical vibration in rated load and different speeds is presented. As it can be found that the performance of this method in 40% and 80% is proper but for other states is not appropriate. Therefore, using this method for detecting fault for this sensor is not possible. As seen in Table 5-9, the fault diagnosis for 100% speed is similar to the performance of this method in Table 5-10 for 100% load. Generally, pre-processing is needed for vibration signals to get a proper result. In Section 5.6, the performance of VAE is presented for different sensors using feature extraction methods, especially for vibration sensors.

Table 5-11: The evaluation of important indexes for 32 classes using the sound signal in CH1 for rated speed and different loads in per unit

Speed	Load	Accuracy	Precision	Recall	F1-Score
100%	No	0.89	0.90	0.90	0.89
	20%	0.82	0.82	0.83	0.81
	40%	0.85	0.87	0.86	0.85
	60%	0.80	0.80	0.81	0.80
	80%	0.68	0.68	0.68	0.68
	100%	0.68	0.66	0.67	0.65
	120%	0.76	0.75	0.77	0.73

Table 5-12: The evaluation of important indexes for 32 classes using the sound signal in CH1 for rated load and different speeds in per unit

Load	Speed	Accuracy	Precision	Recall	F1-Score
100%	20%	-	-	-	-
	40%	0.85	0.88	0.85	0.85
	60%	0.69	0.68	0.71	0.68
	80%	0.94	0.93	0.94	0.93
	100%	0.68	0.66	0.67	0.65
	120%	0.76	0.79	0.77	0.75

Table 5-11 is the evaluation of important indexes results of the sound signal in channel one for rated speed and different loads from no-load to 120% load. As illustrated, the performance of the sound signal is proper, especially for a no-load situation. However, when the load rises, the performance of VAE decreases. The average of accuracy, precision, recall, and F1-Score is 78.28%, 78.28%, 78.85%, and 77.72%, respectively. Table 5-12 is the important indexes results of the sound signal in channel one for rated load and different speeds from 20% to 120%. This table shows that the performance of network at 60% and 100% speeds is not appropriate. Besides, the situation in high speed is not proper same as for high load in Table 5-11. In all of the discussed tables for the important indexes, it can be seen that sometimes precision or other index values are higher than accuracy; the reason for this goes back to their definition in Section 5-2. For instance, when the precision is higher than accuracy, it means that the network's ability to detect faults among all of the predicted faults is better than the network's ability to detect faults and the normal state among all of the tested data.

Table 5-13: The evaluation of important indexes for 32 classes using the sound signal in CH2 for rated speed and different loads in per unit

Speed	Load	Accuracy	Precision	Recall	F1-Score
100%	No	0.92	0.92	0.92	0.92
	20%	0.82	0.83	0.84	0.82
	40%	0.81	0.82	0.82	0.81
	60%	0.84	0.87	0.85	0.84
	80%	0.65	0.65	0.65	0.64
	100%	0.75	0.76	0.75	0.75
	120%	0.76	0.78	0.76	0.75

Table 5-13 presents the important indexes' results of the sound signal in channel two for rated speed and different loads from no-load to 120% load. The average of accuracy, precision, recall, and F1-Score are 79.28%, 80.42%, 79.85%, and 79%, respectively. As shown, the situation for channel two is approximately similar to channel one for the no-load signal. However, the average of important indexes for channel two is better than channel one. It should be noted that the performance of VAE for the sound signal is better than current and vibration signals without extracting features for different loads in rated speed. It shows that feature extraction causes the removal of important information in the sound signals in this condition. Besides, network performance is reduced with increasing load in both channels.

Table 5-14: The evaluation of important indexes for 32 classes using the sound signal in CH2 for rated load and different speeds in per unit

Load	Speed	Accuracy	Precision	Recall	F1-Score
100%	20%	-	-	-	-
	40%	0.94	0.94	0.93	0.93
	60%	0.72	0.72	0.73	0.70
	80%	0.91	0.88	0.91	0.89
	100%	0.64	0.64	0.64	0.63
	120%	0.75	0.76	0.75	0.75

Table 5-14 presents the evaluation results of important indexes for sound signal in channel two for different speeds in rated load. This table shows that the performance is proper compared to the rated speed. Same as the conclusion for Table 5-13, the sound signal at different speeds has better performance than current and vibration signals. In conclusion of this section, it is found that it is needed to extract features from the data because the performance of all of the signals are not proper without feature extraction. Therefore, the VAE is evaluated using feature extraction method in the following of this chapter.

5.5 Variational Autoencoder Analysis with Feature Extraction

This section presents the results of implementing feature extraction methods on VAE such as PCA and LDA as pre-processing methods, described in Chapter 3. This section of this chapter is divided into three crucial sections; analyzing data for separate faults, analyzing data for 32 classes fault, and analyzing for combination of selected faults.

5.5.1 VAE Analysis with Feature Extraction in Rated Condition

The extracted results of important indexes for different faults and different sensors in rated speed and load are prepared in this section. As shown in Table 5-15, a comparison for the current sensor between LDA, PCA, and without feature extraction is presented. The results are evaluated using important indexes for demagnetization, bearing, and eccentricity faults. As shown in Table 5-15, for all faults, LDA has the best stage compared to other methods. Accuracy, precision, recall, and F1-score, for LDA in the current sensor, are 99%, 99%, 99%, and 99%, respectively. However, the PCA situation for detecting different faults is proper. It can be noted that LDA and PCA show excellent performance in detecting bearing and eccentricity faults.

Table 5-15: The indexes using the current signal for rated speed and rated load in per unit

Fault	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Demagnetization	No	0.83	0.84	0.84	0.82
	LDA	1	1	1	1
	PCA	0.99	0.99	0.99	0.99
Bearing	No	0.59	0.62	0.59	0.59
	LDA	0.99	0.99	0.99	0.99
	PCA	0.88	0.88	0.88	0.87
Eccentricity	No	0.62	0.82	0.62	0.64
	LDA	0.98	0.98	0.98	0.98
	PCA	0.97	0.97	0.97	0.97

Table 5-16: The indexes using the horizontal vibration signal for rated speed and rated load in per unit

Fault	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Demagnetization	No	**0.06**	**0**	**0.08**	**0.01**
	LDA	1	1	1	1
	PCA	0.91	0.85	0.92	0.88
Bearing	No	**0.05**	**0**	**0.06**	**0.01**
	LDA	0.99	0.99	0.99	0.99
	PCA	0.91	0.87	0.94	0.90
Eccentricity	No	0.86	0.90	0.87	0.86
	LDA	0.97	0.97	0.97	0.97
	PCA	1	1	1	1

The performance of VAE for horizontal vibration is shown in Table 5-16. The results show that LDA has better performance than other methods for all faults. It can be noted that LDA and PCA in detecting demagnetization and bearing show positive performance. For instance, the accuracy of VAE using LDA for demagnetization is increased from 6% to 100%. Besides, the performance of PCA is better than LDA for eccentricity fault in this investigation. The next sensor is the vertical vibration sensor whose ability for detecting is evaluated as shown in Table 5-17. As presented, PCA has a proper performance compared to LDA.

Table 5-17: The indexes using the vertical vibration signal for rated speed and rated load in per unit

Fault	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Demagnetization	No	0.58	0.52	0.62	0.52
	LDA	0.92	0.87	0.92	0.89
	PCA	1	1	1	1
Bearing	No	0.90	0.92	0.90	0.89
	LDA	1	1	1	1
	PCA	0.99	0.99	0.99	0.99
Eccentricity	No	0.72	0.81	0.76	0.71
	LDA	1	1	1	1
	PCA	1	1	1	1

The performance of the VAE without using any feature extraction methods in vertical direction is better than horizontal direction for demagnetization and bearing faults. It shows that vibration in the vertical direction gets better information for classifying than in the horizontal direction in this case. It should be noted that vibration signals in one direction cannot show all of the important characteristics of faults and it is needed to investigate in different directions.

Table 5-18: The indexes using Sound CH1 signal (for rated speed and rated load) in per unit

Fault	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Demagnetization	No	0.95	0.96	0.96	0.95
	LDA	1	1	1	1
	PCA	1	1	1	1
Bearing	No	0.80	0.81	0.80	0.80
	LDA	1	1	1	1
	PCA	0.99	0.99	0.99	0.99
Eccentricity	No	0.91	0.91	0.91	0.91
	LDA	0.99	0.99	0.99	0.99
	PCA	1	1	1	1

The performance of LDA and PCA are compared for the sound signals in the following. As shown in Table 5-18, PCA and LDA have equal stages compared to no reducing dimensionality for channel one signal. However, the feature extraction method causes to increase in the ability of VAE. As can be seen in Table 5-19, LDA has a better performance compared to other methods. As it can be found from the extracted results in this section, VAE performance is increased considerably using the different pre-processing methods. One of the important results that can be achieved from this section is that the performance of the sound signals for both channels without reducing dimensionality is appropriate in contrast to vibration signals. Besides, LDA and PCA have an effective impact on the performance of VAE for sound signals similar to other signals as well. In the following, the results of implementing VAE for 32 classes are presented consisting of demagnetization, bearing, and eccentricity. Table 5-20 shows the evaluation of results of important indexes for 32 classes in rated speed and rated load with and without feature extraction. It is better to interpret this table using each sensor one by one.

Table 5-19: The indexes using Sound CH2 signal (for rated speed and rated load) in per unit

Fault	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Demagnetization	No	0.99	0.99	0.99	0.99
	LDA	1	1	1	1
	PCA	1	1	1	1
Bearing	No	0.82	0.84	0.84	0.83
	LDA	1	1	1	1
	PCA	0.97	0.97	0.98	0.97
Eccentricity	No	0.97	0.97	0.95	0.96
	LDA	1	1	1	1
	PCA	0.97	0.97	0.97	0.97

As seen, the classification accuracy of the VAE using the current sensor is 65%, and it is increased to 93% for LDA and 88% for PCA. Also, there is the same trend for vibration signals in both directions, and the performance of the VAE is increased to 86% and 99% for the horizontal and vertical direction, respectively. As shown in Table 5-20 for sound signals in channels one and two, similar to the current and vibration signals, the performance of the VAE increases using PCA and LDA. The important indexes of sound signal channel one using pre-processing of LDA are 97%, 97%, and 97 % for precision, recall, and F1-score, respectively. However, this change for sound signals in channels one and two are not as impressive as those for vibration signals.

Table 5-20: The indexes for rated speed and rated load for 32 classes in per unit

Sensor	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Current	No	0.65	0.66	0.64	0.63
	LDA	0.93	0.93	0.93	0.93
	PCA	0.88	0.89	0.89	0.88
Horizontal Vibration	No	0.47	0.41	0.50	0.40
	LDA	0.96	0.96	0.96	0.96
	PCA	0.86	0.79	0.87	0.82
Vertical Vibration	No	0.55	0.54	0.57	0.53
	LDA	0.99	0.99	0.99	0.99
	PCA	0.99	0.99	0.99	0.99
Sound CH 1	No	0.87	0.87	0.87	0.87
	LDA	0.97	0.97	0.97	0.97

	PCA	0.97	0.97	0.97	0.97
Sound CH 2	No	0.84	0.84	0.85	0.84
	LDA	0.97	0.97	0.97	0.97
	PCA	0.98	0.98	0.98	0.98

To conclude, feature extraction has a proper impact on the performance of the VAE for all signals in rated load and rated speed. Also, feature extraction is not essential for sound signals for the low number of classes, though it is better to use the feature extraction methods for the high number of classes. However, using a pre-processing method such as PCA and LDA affects important algorithm indexes such as accuracy and VAE training time discussed in Section 5.7.

Table 5-21: Average of evaluation of important indexes for all of speed and load for different signal in per unit for 32 classes

Sensor	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Horizontal Vibration	No	0.85	0.83	0.83	0.83
	LDA	0.93	0.91	0.92	0.91
	PCA	0.97	0.96	0.96	0.95
Vertical Vibration	No	0.83	0.83	0.83	0.80
	LDA	0.93	0.93	0.93	0.93
	PCA	0.96	0.96	0.96	0.96
Sound CH 1	No	0.84	0.85	0.84	0.84
	LDA	0.92	0.90	0.90	0.89
	PCA	0.97	0.97	0.96	0.95
Sound CH 2	No	0.83	0.84	0.83	0.83
	LDA	0.93	0.91	0.92	0.92
	PCA	0.96	0.94	0.94	0.93

5.5.2 VAE Analysis with Feature Extraction in Different Conditions

In this section, the performance of VAE in all ranges of speeds (20% to120%) and loads (no load to120% load) are evaluated. As shown in Figure 5-22, the average of accuracy, precision, recall, and F1-score for the current signal are 84%, 86%, 84%, and 84%, respectively. The average accuracy is enhanced using feature extraction methods such as PCA and LDA. The accuracy is increased from 84% to 93% for LDA and 95% for PCA, respectively. In contrast to rated speed and rated load, where LDA arrives at a better enhancement in the evaluation of the VAE, PCA has better performance in this section. Table 5-21 presents the performance of VAE using others signals. As it is clear the performance of VAE using PCA and LDA increases for all sensors. The performance of PCA is better than LDA in all cases. Besides, the performance of VAE without any pre-processing in rated condition based on Table 5-20 is not proper; however, its average performance in all condition is proper. Therefore, the VAE investigation in all speed and loads gives better perspective about the its performance.

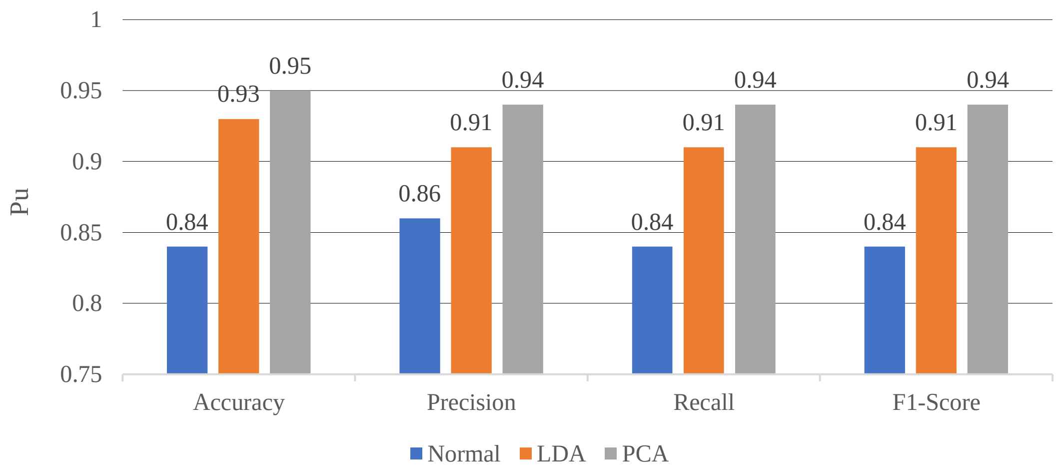


Figure 5-22: Average of evaluation of important indexes for all applied speeds and loads for the current signal in per unit for 32 classes

To conclude, it is found that PCA and LDA are significant to implement in the case of vibration and current signals. Also, PCA shows better performance than LDA as a feature extraction method for classifying data with a high number of classes. Besides, it is found that investigating on rated load and rated speed cannot show the ability of one method completely. As seen, the performance of VAE for all of the speeds and loads is better than rated speed and rated load; however, they are enhanced using pre-processing methods.

5.5.3 VAE Analysis for Combination of Selected Faults

This section investigates the diagnosis of different combinations of faults by combining two, three and all of the faults. As it is impossible to investigate different combinations of all faults and all of the speeds and loads, some cases are selected to investigate in this section, as shown in Table 5-22. The primary aims of this section are to find the ability of VAE for different combinations of faults. It can be noted that the performance of VAE are evaluated with and without feature extraction in this section similar to the previous section. Therefore, this section's structure is divided into three subsections; Analysing VAE for combinations of two faults, three faults, and all studied faults (96 classes).

Table 5-22: Different types of fault for a different combination

Fault	Type of fault			
Demagnetization	10-20-5	50-20-1	50-20-5	50-5-5
Bearing	Ball (1-0.1)	Ball (2-0.15)	Inner (1-0.1)	Outer (1-0.1)
Eccentricity	Ecc-01	Ecc-02		

5.5.3.1 VAE Analysis for Combination of Two Faults

There are 33 classes in this section with different combinations of selected faults in Table 5-22 for analysis. This section presents the results of implementing the VAE in different combinations of two faults. Besides, the detailed results for each class and each sensor are presented in the appendix.

Table 5-23: Average of evaluation of important indexes for different signals for rated speed and load for combinations of two faults in per unit

Sensor	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Current	No	0.71	0.73	0.72	0.70
	LDA	0.95	0.95	0.95	0.95
	PCA	0.98	0.98	0.98	0.98
Horizontal Vibration	No	0.02	0	0.03	0
	LDA	0.99	0.99	0.99	0.99
	PCA	0.96	0.94	0.96	0.95
Vertical Vibration	No	0.28	0.2	0.3	0.21
	LDA	0.99	0.99	0.99	0.99
	PCA	0.95	0.92	0.95	0.94
Sound CH 1	No	0.7	0.7	0.69	0.68
	LDA	0.96	0.96	0.96	0.96
	PCA	0.99	0.99	0.99	0.99
Sound CH 2	No	0.61	0.63	0.62	0.61
	LDA	0.97	0.98	0.98	0.97
	PCA	0.99	0.99	0.99	0.99

As shown in Table 5-23, the average of important indexes for all sensors in the rated condition is presented. As shown, the performance of VAE is increased for all sensors using pre-processing methods. However, the performance of VAE for the current sensor is proper without reducing dimensionality. It can be noted that the PCA in this analysis for the current sensor has better performance in comparison to LDA in the last section. Besides, the complexity of data is increased because of the low average accuracy for this condition compared to single faults for 32 classes. The result for the current sensor using LDA and PCA are 95% and 98% for all indexes, respectively.

Table 5-24: Average of evaluation of important indexes for different signals for rated speed and load for combinations of three faults in per unit

Sensor	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Current	No	0.86	0.87	0.86	0.86
	LDA	0.94	0.93	0.93	0.93
	PCA	0.97	0.98	0.98	0.97
Horizontal Vibration	No	0.02	0	0.03	0
	LDA	0.97	0.97	0.97	0.97
	PCA	0.97	0.94	0.97	0.95
Vertical Vibration	No	0.01	0	0.03	0
	LDA	0.98	0.98	0.98	0.98
	PCA	0.93	0.93	0.93	0.93
Sound CH 1	No	0.61	0.62	0.63	0.61
	LDA	0.94	0.95	0.95	0.94
	PCA	0.98	0.98	0.98	0.97
Sound CH 2	No	0.64	0.64	0.64	0.63
	LDA	0.95	0.95	0.95	0.95
	PCA	0.98	0.98	0.98	0.98

As shown in Table 5-23, the performance of VAE in detecting faults using the vibration signal is not proper. Therefore, using pre-processing method is vital for vibration signals. Besides, the performance of VAE without pre-processing for sound signal is similar to the current signal. However, pre-processing methods increases the performance of VAE. The detailed results of important indexes for different combinations are presented in the appendix. As shown in these tables, the performance of different feature extraction methods is different, and it is not possible to find which one is better for detecting in different combinations.

5.5.3.2 VAE Analysis for Combination of Three Faults

In this section, the ability of VAE for combinations of three faults is investigated. The average results of important indexes are prepared, and the detailed results are prepared in the appendix for each combination of fault. Table 5-24 shows the performance of VAE with and without pre-processing method. The performance of VAE in analysis combination of two faults is similar to analysis for combination of three faults.

According to red highlighted results in Table 5-23 for horizontal and vertical vibration, it is found that VAE without feature extraction is not successful in classifying various classes. To conclude, the PCA for the current and sound signal is the best feature extraction method for combinations of two and three faults. However, the performance of LDA and PCA for vibration signals is different. It can be found that LDA is better than PCA for vibration signals in combinations of two and three faults; however, PCA is better than LDA for one type of fault with different severities.

5.5.3.3 VAE Analysis for Combination of All Faults

In this section, the performance of VAE for all of the combinations consisting of single faults, and combinations of two and three faults are presented. It can be noted the total number of classes in this case is 96. Table 5-25 shows the average of important indexes for VAE with and without pre-processing for different signals. As shown, the performance of VAE is increased using pre-processing methods. The performance of VAE with PCA for all signals is better than VAE with LDA. Besides, the performance of VAE without feature extraction is not proper for horizontal and vertical vibration similar to combination of two and three faults in pervious sections. It shows that when the number of classes rises, the complexity of data increases intensively for vibration signals, and VAE without pre-processing methods cannot classify data.

5.6 Variational Autoencoder Analysis for Mixing Sensors

This section is prepared to evaluate the ability of VAE to diagnose of faults using different combinations of sensors. It can be worthful to describe in detail about the structure of VAE for mixing sensor. The input data is derived from different sensors, and a combination of these sensors was used as input to the VAE model. Prior to feeding the input data into the VAE, a pre-processing step is carried out to ensure that the data is properly formatted. Specifically, at each timestamp, the input data is structured such that there is an array containing different features, one for each sensor. This pre-processing step enabled the VAE to effectively learn the relationship between the different sensors in the input data. The encoder network receives the input data and compresses it into a lower-dimensional latent space representation, which is then passed to the decoder network to reconstruct the original input data. By using the combination of different sensors in the input data, the VAE model is able to capture and learn from the complex relationship between these sensors. Furthermore, the VAE's ability to learn a compressed representation of the data allowed for the efficient capture of complex patterns and relationships in the data. Overall, the use of a VAE model with input from multiple sensors provides a powerful tool for capturing and analyzing complex, multi-dimensional data. The performance evaluation of VAE in this section are prepared in rated speed and rated load for 32 classes of single faults. Besides, the effect of pre-processing methods is investigated on different combinations of sensors.

Table 5-25: Average of evaluation of important indexes for different signals for rated speed and load for combinations of all faults in per unit

Sensor	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Current	No	0.65	0.66	0.65	0.62
	LDA	0.68	0.68	0.68	0.67
	PCA	0.96	0.96	0.96	0.95
Horizontal Vibration	No	0.01	0	0.01	0.01
	LDA	0.81	0.81	0.82	0.81
	PCA	0.91	0.85	0.91	0.88
Vertical Vibration	No	0.01	0	0.01	0
	LDA	0.84	0.84	0.84	0.84
	PCA	0.90	0.88	0.90	0.89
Sound CH 1	No	0.53	0.53	0.53	0.51
	LDA	0.80	0.81	0.80	0.80
	PCA	0.96	0.96	0.97	0.96
Sound CH 2	No	0.53	0.54	0.54	0.51
	LDA	0.81	0.81	0.81	0.80
	PCA	0.96	0.95	0.96	0.95

5.6.1 VAE Analysis for Mixing Two Sensors

This section shows the performance of each combination of two sensors in four tables. Table 5-26 shows the current sensor's performance in combining other sensors. As shown, the combination of the current and other signals without feature extracting is better than the performance of the VAE for the single current or vibration sensors, as shown in Table 5-20. Besides, the performance for the combination of the current and horizontal vibration increases to 95%, and for the combination of the current and vertical vibration rises to 99% without using the pre-processing methods. Also, this situation happens for combinations of current and sound signals. The performance increases from 87% and 84%, respectively, to 94% for both channels.

Table 5-26: Evaluating different combinations of the current and other sensors in rated speed and load in per unit

Sensor 1	Sensor 2	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Current	Horizontal Vibration	No	0.95	0.95	0.94	0.94
		LDA	1	1	1	1
		PCA	1	1	1	1
	Vertical Vibration	No	0.99	0.99	0.98	0.99
		LDA	0.91	0.84	0.91	0.87
		PCA	1	1	1	1
	Sound CH1	No	0.94	0.95	0.95	0.94
		LDA	1	1	1	1
		PCA	0.99	0.99	0.98	0.98
	Sound CH2	No	0.94	0.95	0.95	0.94
		LDA	0.97	0.99	0.97	0.95
		PCA	0.99	0.99	0.99	0.99

It is clear that the combinations of current with other signals directly affect the performance of VAE without using any pre-processing methods. Besides, as seen in Table 5-26, the performance of each combination enhances using LDA and PCA. However, the combinations of current and vertical vibration are reduced using LDA. Moreover, the performance of PCA in all combinations is proper, and the performance of VAE increases for all combinations with pre-processing compared to Table 5-20. The performance of VAE based on extracted results for important indexes for different combinations of horizontal vibration and three sensors are prepared in Table 5-27. Compared to a single sensor, as mentioned in Table 5-20, the performance of VAE increases from 47% and 55%, respectively, to 65% for combinations of both direction vibration signals. Besides, the performance of VAE for this combination is increased to 97% for the accuracy index. The performance of sound signals with horizontal vibration is not proper compared to the single sensor in Table 5-20.

Table 5-27: Evaluating different combination of the current and other sensors in rated speed and load in per unit

Sensor 1	Sensor 2	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Horizontal Vibration	Vertical Vibration	No	0.65	0.62	0.66	0.60
		LDA	0.93	0.89	0.94	0.91
		PCA	0.97	0.94	0.97	0.95
	Sound CH1	No	0.80	0.80	0.81	0.80
		LDA	1	1	1	1
		PCA	0.99	0.99	0.99	0.99
	Sound CH2	No	0.79	0.80	0.79	0.78
		LDA	0.89	0.85	0.91	0.87
		PCA	0.99	0.99	0.99	0.99

Therefore, these combinations cannot be recommended as a helpful sensor data fusion in fault diagnosis of the considered faults. The performance of the data fusion for vertical vibration and the sound sensors in channels one and two, as shown in Table 5-28, are not proper compared to the single sensor in Table 5-20. However, the performance rises using pre-processing methods, especially PCA. It is found that the combinations of vibration and sound signals for VAE are not efficient; however, the performance of VAE with PCA and LDA is improved.

Table 5-28: Evaluating the combination of the vertical vibration and the Sound signals in rated speed and load in per unit

Sensor 1	Sensor 2	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Vertical Vibration	Sound CH1	No	0.78	0.78	0.78	0.77
		LDA	0.94	0.89	0.92	0.90
		PCA	0.99	0.99	0.99	0.99
	Sound CH2	No	0.81	0.82	0.82	0.81
		LDA	1	1	1	1
		PCA	0.99	0.99	0.99	0.99

As shown in Table 5-29, the evaluating results for combinations of sound signals are presented with and without feature extraction. The results show that the performance of sound signals for the combination of channels one and two is reduced from 87% and 84%, respectively, to 69%. Therefore, mixing sound sensors increases the complexity of data to detect. As an important conclusion of the first part of this section, the combination of current signal with other signals without feature extraction is beneficial and causes to increase in the performance of VAE. However, the other combinations' performance for the two sensors without pre-processing is entirely improper.

Table 5-29: Evaluating the combination of the vertical vibration and the Sound signals in rated speed and load in per unit

Sensor 1	Sensor 2	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Sound CH 1	Sound CH 2	No	0.69	0.69	0.70	0.68
		LDA	1	1	1	1
		PCA	0.98	0.98	0.98	0.98

5.6.2 VAE Analysis for Mixing Three Sensors

The performance of VAE for mixing three signals in rated speed and load is shown in Table 5-30. The performance of the current sensor increases the same as for the combinations of two sensors, as shown in Table 5-26. Besides, the performance of the VAE rises to around 100% using PCA.

As seen, the performance of combining vibration signals with other signals has better performance than other combinations of two signals in Table 5-26. It is an important point to note that the performance of most of the states in Table 5-30 reduces for LDA. However, the performance of LDA is better than PCA in the combination of horizontal and vertical vibration with sound channels one and two. It is concluded that the performance of combining three sensors is proper for all states; however, the size of input data using three sensors increases the training time for the proposed method.

Table 5-30: Evaluating the combination of three signals in rated speed and load in per unit

Sensor 1	Sensor 2	Sensor3	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Current	Horizontal vibration	Vertical vibration	No	0.97	0.97	0.97	0.97
			LDA	0.91	0.85	0.91	0.87
			PCA	1	1	1	1
		Sound CH1	No	0.96	0.97	0.96	0.96
			LDA	0.93	0.92	0.94	0.92
			PCA	1	1	1	1
		Sound CH2	No	0.99	0.98	0.99	0.98
			LDA	0.97	0.94	0.97	0.95
			PCA	0.99	0.99	0.99	0.99
	Vertical vibration	Sound CH1	No	0.99	0.99	0.99	0.99
			LDA	0.92	0.86	0.91	0.88
			PCA	1	1	1	1
		Sound CH2	No	0.97	0.97	0.97	0.97
			LDA	0.83	0.78	0.84	0.80
			PCA	0.99	0.99	0.99	0.99
	Sound CH1	Sound CH2	No	0.90	0.91	0.91	0.90
			LDA	0.87	0.80	0.87	0.82
			PCA	0.99	0.99	0.99	0.99
Horizontal vibration	Vertical vibration	Sound CH1	No	0.96	0.97	0.96	0.96
			LDA	0.94	0.90	0.91	0.91
			PCA	1	1	0.99	1
		Sound CH2	**No**	0.97	0.98	0.97	0.97
			LDA	0.92	0.89	0.94	0.91
			PCA	0.96	0.94	0.97	0.95
	Sound CH1	Sound CH2	No	0.79	0.79	0.8	0.79
			LDA	0.91	0.84	0.91	0.87
			PCA	0.9	0.98	0.99	0.99
Vertical vibration	Sound CH1	Sound CH2	No	0.81	0.82	0.81	0.81
			LDA	1	1	1	1
			PCA	0.99	0.99	0.99	0.99

5.6.3 VAE Analysis for Mixing Four Sensors

Table 5-31 shows the performance of the combination of four sensors in rated speed and rated load with and without feature extraction. As shown, the performance of all states is proper; however, increasing the number of sensors raises the training time. It is an important point to note in this table that the performance of the first combination is reduced to zero using LDA. Therefore, LDA is not helpful in all cases. Also, VAE with PCA is an effective method based on Table 5-31 similar to single sensor and a combination of two sensors.

Table 5-31: Evaluating the combination of four sensor signals in rated speed and load in per unit

Sensor 1	Sensor 2	Sensor3	Sensor 4	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Current	Horizontal vibration	Vertical Vibration	Sound CH1	**No**	1	1	1	1
				LDA	0.04	0	0.03	0
				PCA	1	1	1	1
			Sound CH2	No	0.99	0.99	0.99	0.99
				LDA	0.69	0.53	0.69	0.58
				PCA	1	1	1	1
		Sound CH1	Sound CH2	No	0.98	0.98	0.98	0.98
				LDA	0.87	0.80	0.88	0.82
				PCA	0.99	0.99	0.99	0.99
	Vertical vibration	Sound CH1	Sound CH2	No	0.98	0.98	0.98	0.98
				LDA	0.76	0.66	0.76	0.70
				PCA	0.99	0.99	0.99	0.99
Horizontal vibration	Vertical vibration	Sound CH1	Sound CH2	**No**	0.96	0.97	0.96	0.96
				LDA	0.84	0.74	0.84	0.78
				PCA	0.96	0.93	0.96	0.94

5.6.4 VAE Analysis for Mixing Five Sensors

As shown in Table 5-32, the performance of all of the sensors together is presented. As it can be deduced from the results of this table, the performance of VAE without pre-processing is proper, and the performance of VAE with pre-processing method is reduced. Therefore, it is found that when all of the extracted data in this thesis are combined, the pre-processing method has a not proper effect on VAE.

Table 5-32: Evaluating the combination of five signals in rated speed and load in per unit

Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Reducing dimensionality	Accuracy	Precision	Recall	F1-Score
Current	Horizontal vibration	Vertical vibration	Sound CH1	Sound CH2	**No**	0.99	0.99	0.99	0.99
					LDA	0.87	0.79	0.88	0.82
					PCA	0.98	0.98	0.98	0.98

To conclude, different results are achieved by combining different sensors' data, and it is impossible to get a general conclusion about different combinations of sensors' data. However, the combination of current and other signals using PCA as a pre-processing method is slightly better than others from accuracy and training time perspectives. Besides, it is found that the pre-processing method cannot enhance the performance of VAE in all cases of combining sensors, especially for implementing LDA for combining three, four, and five sensors.

5.7 Comparison Between VAE and Machine Learning Methods

In this section, the comparison is made between VAE and machine learning methods consisting of NB, DT, and KNN to show the ability of VAE in aspects of important indexes and training time. Training time is an essential index for implementing the new method on hardware. However, the evaluation of this index is prepared for comparison between VAE, with and without feature extraction and machine learning methods in this section. This comparison is made for two series of data consisting of 32 and 65 data classes for each sensor in rated load and speed.

Table 5-33: Comparison between VAE and machine learning for the Current rated speed and load for 32 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	061	0.60	0.52	0.58	0:00:59.91
	LDA	**0.94**	0.94	0.93	0.94	**0:00:52.23**
	PCA	0.89	0.89	0.89	0.88	0:00:51.52
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:01:32**
	NB	0.03	0.01	0.03	0.01	0:00:0.7
	DT	0.94	0.94	0.94	0.94	0:40:09

5.7.1 Comparing between VAE and Machine Learning for 32 Classes

As shown in Table 5-33, the evaluation of different indexes for the current sensor is presented. Generally, it is found that the training times of VAE with and without pre-processing methods are not so different. However, the performance of machine learning methods in the training time index except NB is proper. Despite the stochastic independence of the features, the performance of the naive Bayes algorithm is inadequate in such cases. It can be noted that VAE with LDA has the best performance among different methods with 94% accuracy and 0:00:52.23 training time. Besides, the training time of DT is approximately 0:40:09, which is an improper training time for a diagnostic algorithm.

Table 5-34: Comparison between VAE and machine learning for the horizontal vibration in rated speed and load for 32 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.61	0.63	0.62	0.59	0:01:00.17
	LDA	**0.93**	0.93	0.93	0.93	**0:00:52.88**
	PCA	0.91	0.91	0.91	0.91	0:00:53.35
Machine Learning	**KNN**	**0.99**	0.99	0.99	0.99	**0:02:28.53**
	NB	0.15	0.08	0.15	0.09	0:00:00.79
	DT	0.98	0.98	0.98	0.98	0:01:24.82

The performance of VAE for vibration signals in horizontal and vertical directions are presented in Table 5-34 and 5-35, respectively. The performance of KNN and DT are 99% and 98% accuracy with 0:02:28.53 and 0:01:24.82 training time for horizontal vibration as shown in Table 5-34, respectively. Besides, the performance of VAE with PCA and LDA is proper for accuracy and training time together. As seen, the accuracy of VAE for LDA and PCA are 93% and 91% for LDA and PCA with 0:00:52.88 and 0:00:53.35, respectively.

Table 5-35: Comparison between VAE and machine learning for the Vertical Vibration in rated speed and load for 32 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.64	0.65	0.65	0.61	0:00:58.80
	LDA	**0.94**	0.94	0.94	0.94	**0:00:54.19**
	PCA	0.90	0.91	0.91	0.90	0:00:59.17
Machine Learning	**KNN**	**0.99**	0.99	0.99	0.99	**0:02:20.99**
	NB	0.13	0.60	0.13	0.06	0:00:00.86
	DT	0.98	0.98	0.98	0.98	0:01:25.11

Table 5-35 shows that vertical vibration's performance which is approximately the same as horizontal vibration. It can be noted that the performance of DT for the training time index is better for the vibration signal compared to the current signal. Besides, the performance of VAE for sound signals in channel one is approximately the same as for previous signals as presented in Table 5-36.

Table 5-36: Comparison between VAE and machine learning for the Sound CH1 in rated speed and load for 32 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.62	0.63	0.63	0.60	0:01:01.23
	LDA	**0.93**	0.93	0.93	0.93	**0:00:56.42**
	PCA	0.91	0.91	0.90	0.90	0:00:58.27
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:01:49.27**
	NB	0.05	0.01	0.05	0.02	0:00:00.79
	DT	0.92	0.92	0.92	0.92	0:22:54.41

The performance of VAE and machine learning methods for sound signals in channel two is presented in Table 5-37. The performance of PCA in channel two is not the same as in channel one; the performance of VAE using LDA in channel two with 93% accuracy and 0:00:58.74 for training time is better than PCA, and approximately the same as channel one. Besides, it is concluded that DT's performance depends on the signal type because the training time for different signals show different performance.

Table 5-37: Comparison between VAE and machine learning for the Sound CH2 in rated speed and load for 32 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.61	0.62	0.60	0.59	0:01:01.89
	LDA	**0.93**	0.93	0.93	0.93	**0:00:58.74**
	PCA	0.88	0.88	0.88	0.88	0:00:57.55
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:01:46.15**
	NB	0.05	0.01	0.05	0.02	0:00:00.79
	DT	0.92	0.92	0.92	0.92	0:21:17.23

To conclude, it is found that the performance of machine learning method in the aspect of the important indexes such as precision are better than VAE; however, the training time of VAE using feature extraction methods are proper. Therefore, a trade-off between important indexes and training time is needed based on each case study. It can be helpful to evaluate the performance of VAE and machine learning methods for a high number of classes. In the following, the performance of VAE and machine learning method for different signals are benchmarked using 65 data classes consisting of normal and a combination of different faults which described in Section 5.4.3.

Table 5-38: Comparison between VAE and machine learning for the current signal in rated speed and load for 65 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.52	0.51	0.54	0.50	0:01:11.18
	LDA	0.82	0.83	0.82	0.82	0:01:07.78
	PCA	**0.97**	0.97	0.97	0.97	**0:01:07.04**
Machine Learning	**KNN**	**0.90**	0.90	0.90	0.90	**0:07:03.75**
	NB	0.02	0	0.02	0	0:00:03.05
	DT	0.73	0.73	0.73	0.73	1:38:16.77

5.7.2 Comparing between VAE and Machine Learning for 65 Classes

Table 5-38 shows the performance of VAE and machine learning for the current signal based on important indexes and training time. The performance of PCA is better than LDA and machine learning methods, for both aspects. The LDA performance has a 97% accuracy, and training time is 0:01:07.04. The performance of DT in both aspects is not proper, its classification accuracy is 73%, and training time is 1:38:16.77. It is concluded that the performance of VAE using PCA when the classes number is increased is better than the machine learning method for the current signal.

Table 5-39: Comparison between VAE and machine learning for the horizontal vibration in rated speed and load for 65 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.61	0.62	0.62	0.59	0:01:13.15
	LDA	0.82	0.82	0.83	0.82	0:01:12.32
	PCA	**0.96**	0.96	0.96	0.96	**0:01:10.80**
Machine Learning	**KNN**	**0.97**	0.97	0.97	0.97	**0:09:06.76**
	NB	0.08	0.04	0.08	0.04	0:00:03.01
	DT	0.94	0.94	0.94	0.94	0:04:30.25

Tables 5-39 and 5-40 show the performance of VAE for vibration signal in horizontal and vertical directions. As shown in Table 5-38, the performance of KNN is better than PCA with 97% classification accuracy and 0:09:06.76 training time for horizontal vibration. However, the performance of PCA is better than KNN for the training time aspect with 0:01:10.80. As in the case of the current sensor; when the number of classes increases, the performance of VAE using PCA enhances with better training time.

Table 5-40: Comparison between VAE and machine learning for the vertical vibration in rated speed and load for 65 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.55	0.54	0.53	0.55	0:01:15.45
	LDA	0.81	0.81	0.81	0.81	0:01:14.70
	PCA	**0.97**	0.97	0.97	0.97	**0:01:13.47**
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:12:44.49**
	NB	0.12	0.04	0.12	0.04	0:00:03.01
	DT	0.96	0.96	0.96	0.96	0:03:41.27

As shown in Table 5-40, the performance of vertical vibration is similar to horizontal vibration described in the previous table. The performance of PCA with 97% accuracy and 0:01:13.47 training time is appropriate; however, the performance of KNN is better than PCA in the important index aspect. It can be concluded that for vibration signals, the performance of the VAE using feature extraction is better than machine learning methods for both classification accuracy and training time.

Table 5-41: Comparison between VAE and machine learning for the sound CH1 in rated speed and load for 65 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.60	0.62	0.62	0.60	0:01:19.11
	LDA	0.83	0.84	0.83	0.83	0:01:18.00
	PCA	**0.97**	0.97	0.97	0.97	**0:01:19.03**
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:08:40.15**
	NB	0.03	0	0.03	0	0:00:03.05
	DT	0.89	0.89	0.89	0.89	1:20:45.19

Table 5-41 shows the performance of the VAE and machine learning methods for sound in channel one using important indexes and training time. The performance of KNN is proper in the important indexes aspect with 98% accuracy, and PCA is proper for the training time aspect with 0:01:19.03. Besides, the performance of DT is not proper because of slow training time similar to the current signal, as shown in Table 5-41.

The performance of sound channel two is evaluated as shown in Table 5-42. As it can be found similar to channel one, the performance of KNN is the best for classification accuracy with 98%, and PCA is better than other methods in training time with 0:01:19.24. It can be concluded that, it is needed a trade-off to get better results.

Table 5-42: Comparison between VAE and machine learning for the sound CH2 in rated speed and load for 65 classes in per unit

Sensor	Method	Accuracy	Precision	Recall	F1-Score	Training time
VAE	No	0.57	0.59	0.58	0.58	0:01:20.19
	LDA	0.83	0.84	0.84	0.84	0:01:18.00
	PCA	**0.97**	0.97	0.97	0.97	**0:01:19.24**
Machine Learning	**KNN**	**0.98**	0.98	0.98	0.98	**0:08:32.04**
	NB	0.03	0	0.03	0	0:00:02.94
	DT	0.90	0.90	0.90	0.90	1:20:18.53

To conclude, it is found that the performance of VAE using PCA and LDA in a high number of classes is better than other machine learning methods in both aspects such as training time and the important indexes. However, in the low number of classes, the performance of KNN and DT for the important indexes is better than VAE with and without pre-processing methods. Also, it can be noted that the performance of VAE with pre-processing in low classes of data for training time is better than machine learning methods.

5.8 Summary

This chapter presents the investigation of the proposed method for diagnosing faults in a prototyped PMSM. The performance of the VAE in different conditions with different signals are evaluated, and its ability is developed using the pre-processing method. As it is illustrated the performance of the VAE with PCA in most of the analysing cases is acceptable. Implementing pre-processing methods on vibration signals is one of the essential steps in classifying faults using VAE. Besides, analysing data using vibration signals in both directions can be more successful than in one direction. Compared to other types of signals, sound signals offer a cost-effective and accurate means of classifying data. Besides, the combination of sound and current signals based on the high accuracy in classifying with and without pre-processing methods can be a proper way for enhancing the performance of VAE. However, using more signals for classifying data using VAE increases the network training time.

6 CONCLUSION

This chapter presents the complete conclusion from all of the extracted results and other important notes of this thesis. This chapter is divided into two sections; the conclusion and the suggestion to continue this research work in the future.

6.1 Conclusion

The VAE is investigated using different signals in different loads and speeds, and it is analyzed for different combinations of faults and sensors. Besides, the performance of VAE using feature extraction methods is evaluated, and it is compared to several machine learning methods to show its ability for different indexes and training time. As described in Section 5.1, different indexes comprising of accuracy, precision, recall and F1-score are considered as the important indexes to evaluate the ability of VAE. Besides, the confusion matrix and ROC curve are presented as efficient tools for evaluating the performance of proposed and machine learning methods. The important indexes and confusion matrix are selected as proper tools for evaluating the performance of the proposed method. It can be helpful to present the important conclusions for the VAE in the following of this chapter for different signals, separately. However, presenting all of the VAE ability is not possible in this section, and it is needed to come back to Chapter 5 to investigate these in detail.

The performance of the current sensor for VAE is not satisfactory for all of the cases of faults and speeds and loads range without using the feature extraction methods. It means that VAE did not succeed in extracting important features to diagnose different faults without using pre-processing methods such as PCA and LDA. However, PCA and LDA as feature extraction methods have an excellent effect on the current sensor for different conditions. For instance, the average accuracy of the current sensor for detecting faults in 32 classes is increased from 65% to 93% using LDA in rated speed and rated loads. Also, the accuracy is increased for 96 classes from 65% to 96% using PCA. It is important to note that when the number of classes increases, the performance of PCA is better than LDA as a feature extraction method for the current signal. Besides, the performance of current signals for different combinations of faults is proper using PCA as a feature extraction method, as it is increased from 86% to 97% for combinations of three faults.

The ability of the vibration sensor, one of the most applicable sensors in industrial applications, for both directions is approximately the same. It is found that vibration sensors' performance is enhanced considerably by implementing the pre-processing method.

For instance, the performance of VAE using horizontal vibration for detecting 96 classes data is increased from 47% to 96% using LDA as a pre-processing method. Also, it is increased from 1% to 91% using PCA. The results show that detecting faults using VAE without using pre-processing method is no easy work. Moreover, the ability of VAE for the combination of two faults’ data increases from 2% to 99% using LDA as a feature extraction method. In this sensor, LDA performs better in the low and high number of classes as a pre-processing method.

The performance of sound signals for two channels are analyzed, and interesting results are extracted. In contrast to current and vibration signals, it is found that the VAE has proper performance without using pre-processing. The accuracy of VAE for detecting 32 classes data without any feature extracting method is around 87%, and it enhanced to 97% using LDA. It shows that feature extraction has a proper effect on VAE to detect the fault using vibration. However, the performance of sound signals for combinations of two or three faults is not proper. For instance, the performance of sound signal in channel one for combinations of two faults is around 70%, and it is enhanced to 99% using PCA. The performance of VAE for detecting 96 classes data is around 53% and increased to 96% using PCA. The VAE with LDA in the low number of classes and VAE with PCA in the high number of classes have the best performance in classifying data.

Another investigation is done on VAE to analyze the mixing of two, three, four, and five sensors to achieve a proper perspective about the data fusion. The performance of VAE in mixing two sensors revealed that the effectiveness of combining different signals varies. Therefore, it cannot be definitively concluded that mixing two sensors is always a successful approach for enhancing the VAE's ability. However, it is found that the combination of current with another signal is a successful combination and it causes to enhance the performance of VAE. Besides, the proposed method’s performance is enhanced using feature extraction methods. Mixing three, four, and five sensors shows that mixing sensors increase the chance of getting better results.

However, when the number of sensors is increased, the data size increases, and thus the training time rises. Besides, the large data size needs better hardware to save and process the data. Therefore, mixing sensors without any data fusion methods are possible for VAE; however, the restrictions of RAM and GPU should be noted. It can be noted that the performance of VAE for mixing the sensors increased when feature extraction is implemented on the data though it is not true for some cases and some signals. For Instance, the performance of VAE with LDA is decreased compared to VAE without using pre-processing when the current, horizontal and vertical vibration, and sound signals in channel 1 are mixed. Therefore, VAE with feature extraction cannot be helpful all of the time.

A comparison is prepared between the VAE and machine learning algorithms to show the benefits and drawbacks of this method in relation to important indexes and training time. DT, NB, and KNN are selected as machine learning methods in this comparison. Besides, this research is done for 32 and 65 data classes. It is found that the performance of VAE using PCA and LDA in a high number of classes is better than other machine learning methods in both aspects, such as training time and the important indexes. However, for the low number of classes, the performance of KNN and DT for the important indexes is better than VAE with and without pre-processing methods. Also, it can be noted that the performance of VAE with pre-processing in low data classes for training time is better than machine learning methods.

It is demonstrated that the performance of VAE with PCA as a pre-processing method is increased considerably compared to LDA for most of the case studies. Therefore, VAE with PCA is a sufficient method for classifying data in this thesis. Besides, the combination of current and sound signals is the best among all of them. Therefore, the combination of current and sound signals using VAE with PCA can be a good method for diagnosing faults in this study.

6.2 Future Work

According to the achieved experiences from this research work, some suggestions are prepared in the following. It is found that sound signals compared to other signals, have better performance for other indexes such as accuracy. Also, the performance of VAE is increased for the combination of sound and current signals compared to other combinations. Therefore, this combination for VAE can be developed for different case studies and different conditions. Besides, the sound signal is the cheapest one among all of the sensors in this research work. Therefore, it is recommended to extend working on sound signals for more electrical machines and for different faults in various loads and speeds. Also, it is suggested to work on different pre-processing methods to increase the performance of VAE. A lot of effort was made to prepare the optimized VAE in this thesis; however, the VAE hyperparameters were optimized manually because of the restriction of GPU and memories. Therefore, it can be defined as a new project to find the optimized hyperparameters of VAE.

Another important suggestion is to continue working on deep learning for condition monitoring of the electrical machine. According to this research work, deep learning, compared to machine learning, has better accuracy and training time, especially in high data classes. Besides, Transfer learning can be presented as a novel method in condition monitoring of electrical machines in which the achieved knowledge about one electrical machine can be transferred to another type of electrical machine.

Working on experimental data has an essential role in improving this method and industrializing it. It can be a proper suggestion to continue working on collected data by the industrial company to get a better view of the industrial machines' faults and the ability of the proposed method to be implemented in industrial applications. Therefore, it is strongly recommended to continue working on industrial data instead of working in an experimental setup in a laboratory.

LIST OF FIGURES

LIST OF TABLES

7 APPENDIX

Table A-1: The important index of VAE using the current sensor in rated speed and rated load for demagnetization and eccentricity combination in per unit

Demagnetization	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ecc-01	No	0.90	0.70	0.79
		LDA	0.96	0.93	0.94
		PCA	1	1	1
	Ecc-02	No	0.48	0.86	0.61
		LDA	1	0.93	0.96
		PCA	1	1	1
50-20-1	Ecc-01	No	0.56	0.70	0.62
		LDA	1	0.93	0.96
		PCA	1	1	1
	Ecc-02	No	0.82	0.59	0.69
		LDA	0.84	1	0.91
		PCA	1	1	1
50-20-5	Ecc-01	No	1	1	1
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.7	0.93	0.8
		LDA	1	1	1
		PCA	0.97	0.91	0.94
50-5-5	Ecc-01	No	0.47	0.5	0.48
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.59	0.83	0.69
		LDA	0.94	0.86	0.90
		PCA	1	1	1

Table A-2: The important index of VAE using the current sensor in rated speed and rated load for the bearing and eccentricity combination in per unit

Bearing	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
Ball-101	Ecc-01	No	0.5	0.51	0.50
		LDA	0.96	1	0.98

		PCA	1	1	1
	Ecc-02	No	0.72	0.37	0.49
		LDA	0.90	0.96	0.93
		PCA	0.95	0.77	0.85
Ball-2015	Ecc-01	No	0.56	0.80	0.66
		LDA	0.85	0.88	0.87
		PCA	0.95	1	0.97
	Ecc-02	No	0.79	0.87	0.83
		LDA	1	0.97	0.98
		PCA	1	1	1
Inner-101	Ecc-01	No	0.75	0.87	0.80
		LDA	0.95	0.86	0.90
		PCA	1	1	1
	Ecc-02	No	0.44	0.68	0.54
		LDA	1	0.85	0.92
		PCA	1	1	1
Outer-101	Ecc-01	No	0.79	0.70	0.74
		LDA	1	0.93	0.96
		PCA	1	1	1
	Ecc-02	No	0.35	0.41	0.38
		LDA	0.88	0.82	0.85
		PCA	0.92	0.97	0.94

Table A-3: The important index of VAE using the current sensor in rated speed and rated load for demagnetization and bearing combination in per unit

Demagnetization	Bearing	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ball-101	No	0.96	0.75	0.84
		LDA	0.96	0.96	0.96
		PCA	1	1	1
	Ball-2015	No	0.85	0.60	0.70
		LDA	0.78	1	0.88
		PCA	1	1	1
	Inner-101	No	0.53	0.23	0.32
		LDA	0.96	1	0.98
		PCA	0.89	0.96	0.92
		No	0.34	0.27	0.30

	Outer-101	LDA	0.86	0.96	0.91
		PCA	1	0.96	0.98
50-20-1	Ball-101	No	1	1	1
		LDA	1	1	1
		PCA	1	1	1
	Ball-2015	No	0.87	0.96	0.92
		LDA	1	1	1
		PCA	1	1	1
	Inner-101	No	0.48	0.59	0.53
		LDA	1	0.96	0.98
		PCA	1	1	1
	Outer-101	No	0.88	0.96	0.92
		LDA	0.97	1	0.98
		PCA	1	1	1
50-20-5	Ball-101	No	0.93	0.63	0.75
		LDA	1	1	1
		PCA	0.89	0.96	0.92
	Ball-2015	No	0.97	0.97	0.97
		LDA	0.961	1	0.98
		PCA	1	1	1
	Inner-101	No	1	1	1
		LDA	0.97	1	0.98
		PCA	1	1	1
	Outer-101	No	0.94	0.80	0.87
		LDA	1	1	1
		PCA	1	1	1
50-5-5	Ball-101	No	0.66	0.23	0.34
		LDA	1	1	1
		PCA	0.96	1	0.98
	Ball-2015	No	1	0.89	0.94
		LDA	0.89	1	0.94
		PCA	0.97	0.91	0.94
	Inner-101	No	0.69	0.58	0.63
		LDA	0.84	0.84	0.84

		PCA	1	0.97	0.98
	Outer-101	No	0.86	0.74	0.8
		LDA	1	1	1
		PCA	1	1	1
-	-	No	0.52	0.96	0.68
		LDA	0.91	0.77	0.83
		PCA	0.88	1	0.94

Table A-4: The important index of VAE using the horizontal vibration sensor in rated speed and rated load for demagnetization and eccentricity combination in per unit

Demagnetization	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	0.86	1	0.93
	Ecc-02	No	0	0	0
		LDA	1	0.97	0.98
		PCA	0.91	0.97	0.94
50-20-1	Ecc-01	No	0	0	0
		LDA	0.96	0.96	0.96
		PCA	0.73	0.97	0.84
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
50-20-5	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
50-5-5	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	0.92	1	0.96

Table A-5: The important index of VAE using the horizontal vibration sensor in rated speed and rated load for bearing and eccentricity combination in per unit

Bearing	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
Ball-101	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	0.9	1	0.95
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
Ball-2015	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0	0	0
		LDA	0.93	1	0.97
		PCA	1	1	1
Inner-101	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.98
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
Outer-101	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	1	0.97	0.98
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.99

Table A-6: The important index of VAE using the horizontal vibration sensor in rated speed and rated load for demagnetization and bearing combination in per unit

Demagnetization	Bearing	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ball-101	No	0	0	0
		LDA	0.94	1	0.97
		PCA	0.93	1	0.96
		No	0	0	0

	Ball-2015	LDA	0.97	0.94	0.96
		PCA	0.97	1	0.99
	Inner-101	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.99
50-20-1	Ball-101	No	0.02	1	0.04
		LDA	1	1	1
		PCA	0	0	0
	Ball-2015	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
	Inner-101	No	0	0	0
		LDA	1	0.97	0.98
		PCA	0.94	1	0.97
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
50-20-5	Ball-101	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.99
	Ball-2015	No	0	0	0
		LDA	0.97	0.94	0.96
		PCA	1	1	1
	Inner-101	No	0	0	0
		LDA	1	1	1
		PCA	0.96	1	0.98
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	1	0.97	0.98
50-5-5	Ball-101	No	0	0	0
		LDA	1	1	1

		PCA	0.96	1	0.98
	Ball-2015	No	0	0	0
		LDA	1	1	1
		PCA	1	0.97	0.99
	Inner-101	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.98
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	1	0.97	0.98
-	-	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.98

Table A-7: The important index of VAE using the vertical vibration sensor in rated speed and rated load for demagnetization and eccentricity combination in per unit

Demagnetization	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ecc-01	No	0.26	0.29	0.28
		LDA	1	1	1
		PCA	1	0.96	0.98
	Ecc-02	No	0	0	0
		LDA	1	0.96	0.98
		PCA	0.88	1	0.94
50-20-1	Ecc-01	No	0	0	0
		LDA	0.97	1	0.99
		PCA	0.96	1	0.98
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
50-20-5	Ecc-01	No	0	0	0
		LDA	1	0.93	0.96
		PCA	0.97	0.97	0.97
	Ecc-02	No	0.77	1	0.87
		LDA	0.97	1	0.99
		PCA	0.66	1	0.79
50-5-5	Ecc-01	No	0	0	0

		LDA	0.92	1	0.96
		PCA	0.98	1	0.99
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.98

Table A-8: The important index of VAE using the vertical vibration sensor in rated speed and rated load for the bearing and eccentricity combination in per unit

Bearing	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
Ball-101	Ecc-01	No	0.09	0.04	0.05
		LDA	1	1	1
		PCA	1	0.93	0.96
	Ecc-02	No	0	0	0
		LDA	0.97	1	0.98
		PCA	0.97	0.86	0.91
Ball-2015	Ecc-01	No	0.41	1	0.58
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.12	0.42	0.19
		LDA	1	1	1
		PCA	1	1	1
Inner-101	Ecc-01	No	0	0	0
		LDA	1	1	1
		PCA	0.96	1	0.98
	Ecc-02	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
Outer-101	Ecc-01	No	0.53	1	0.69
		LDA	1	1	1
		PCA	0.97	1	0.98
	Ecc-02	No	0	0	0
		LDA	0.94	1	0.97
		PCA	1	1	1

Table A-9: The important index of VAE using the vertical vibration sensor in rated speed and rated load for demagnetization and bearing combination in per unit

Demagnetization	Bearing	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ball-101	No	0.15	1	0.27
		LDA	1	0.97	0.99
		PCA	0.74	0.8	0.77
	Ball-2015	No	0	0	0
		LDA	1	1	1
		PCA	0.86	1	0.93
	Inner-101	No	0.28	0.29	0.29
		LDA	0.96	1	0.98
		PCA	0.74	0.94	0.83
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	0.97	1	0.99
50-20-1	Ball-101	No	0.36	0.32	0.34
		LDA	0.96	1	0.98
		PCA	1	1	1
	Ball-2015	No	0.53	0.33	0.41
		LDA	1	1	1
		PCA	0.96	1	0.98
	Inner-101	No	0.33	1	0.49
		LDA	1	0.94	0.97
		PCA	1	1	1
	Outer-101	No	0.88	0.41	0.56
		LDA	1	1	1
		PCA	1	1	1
50-20-5	Ball-101	No	0.07	0.04	0.05
		LDA	1	1	1
		PCA	0.94	1	0.97
	Ball-2015	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
		No	1	1	1

	Inner-101	LDA	1	1	1
		PCA	0	0	0
	Outer-101	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1
50-5-5	Ball-101	No	0.54	0.39	0.45
		LDA	1	1	1
		PCA	1	1	1
	Ball-2015	No	0.09	0.84	0.17
		LDA	1	1	1
		PCA	1	1	1
	Inner-101	No	0.42	0.86	0.56
		LDA	1	0.94	0.97
		PCA	1	1	1
	Outer-101	No	0	0	0
		LDA	1	0.94	0.97
		PCA	0.97	1	0.98
-	-	No	0	0	0
		LDA	1	1	1
		PCA	1	1	1

Table A-10: The important index of VAE using the sound sensor CH1 in rated speed and rated load for demagnetization and eccentricity combination in per unit

Demagnetization	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ecc-01	No	0.8	0.7	0.74
		LDA	0.96	1	0.98
		PCA	1	1	1
	Ecc-02	No	0.79	0.51	0.62
		LDA	0.97	1	0.99
		PCA	1	1	1
50-20-1	Ecc-01	No	0.89	0.95	0.92
		LDA	0.97	1	0.99
		PCA	1	1	1
	Ecc-02	No	0.92	0.88	0.9
		LDA	1	1	1

		PCA	1	1	1
50-20-5	Ecc-01	No	0.67	0.54	0.6
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.79	0.84	0.82
		LDA	1	1	1
		PCA	0.96	1	0.98
50-5-5	Ecc-01	No	0.93	0.34	0.5
		LDA	1	0.93	0.96
		PCA	1	1	1
	Ecc-02	No	0.85	0.7	0.77
		LDA	1	1	1
		PCA	1	1	1

Table A-11: The important index of VAE using sound sensor CH1 in rated speed and rated load for bearing and eccentricity combination in per unit

Bearing	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
Ball-101	Ecc-01	No	0.71	0.5	0.59
		LDA	0.96	1	0.98
		PCA	1	1	1
	Ecc-02	No	0.52	0.81	0.63
		LDA	1	0.97	0.98
		PCA	1	0.93	0.96
Ball-2015	Ecc-01	No	0.68	0.59	0.63
		LDA	0.96	0.93	0.95
		PCA	1	1	1
	Ecc-02	No	0.74	0.7	0.72
		LDA	0.96	1	0.98
		PCA	1	1	1
Inner-101	Ecc-01	No	0.35	0.32	0.33
		LDA	1	1	1
		PCA	1	0.97	0.98
	Ecc-02	No	0.65	1	0.79
		LDA	0.97	1	0.98
		PCA	0.9	1	0.95
Outer-101	Ecc-01	No	0.67	0.69	0.68

		LDA	1	0.94	0.97
		PCA	1	1	1
	Ecc-02	No	0.48	0.38	0.42
		LDA	1	1	1
		PCA	1	0.97	0.99

Table A-12: The important index of VAE using sound sensor CH1 in rated speed and rated load for demagnetization and bearing combination in per unit

Demagnetization	Bearing	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ball-101	No	0.56	0.36	0.44
		LDA	0.88	1	0.93
		PCA	1	0.97	0.98
	Ball-2015	No	0.69	0.87	0.77
		LDA	0.97	0.97	0.97
		PCA	1	1	1
	Inner-101	No	0.78	0.91	0.84
		LDA	0.9	1	0.95
		PCA	1	0.84	0.92
	Outer-101	No	0.64	0.9	0.75
		LDA	0.96	0.93	0.95
		PCA	0	0	0
50-20-1	Ball-101	No	0.71	0.85	0.77
		LDA	0.96	0.96	0.96
		PCA	1	1	1
	Ball-2015	No	0.73	0.31	0.43
		LDA	0.97	1	0.98
		PCA	1	0.97	0.98
	Inner-101	No	0.35	0.28	0.31
		LDA	1	1	1
		PCA	1	1	1
	Outer-101	No	0.61	0.52	0.56
		LDA	0.92	0.89	0.91
		PCA	0.97	0.94	0.96
50-20-5	Ball-101	No	0.56	0.78	0.65
		LDA	0.9	1	0.95

		PCA	1	1	1
	Ball-2015	No	0.5	0.69	0.58
		LDA	0.97	0.97	0.97
		PCA	0.93	1	0.96
	Inner-101	No	0.79	0.71	0.75
		LDA	1	0.85	0.92
		PCA	1	1	1
	Outer-101	No	0.51	0.82	0.63
		LDA	1	1	1
		PCA	1	1	1
50-5-5	Ball-101	No	0.75	0.92	0.83
		LDA	0.86	0.91	0.88
		PCA	1	1	1
	Ball-2015	No	1	0.94	0.97
		LDA	0.97	1	0.99
		PCA	1	1	1
	Inner-101	No	0.88	0.77	0.82
		LDA	0.93	0.93	0.93
		PCA	1	1	1
	Outer-101	No	0.78	0.75	0.76
		LDA	0.97	0.81	0.88
		PCA	1	1	1
-	-	No	0.75	0.96	0.84
		LDA	0.88	0.85	0.87
		PCA	1	1	1

Table A-13: The important index of VAE using the sound sensor CH2 in rated speed and rated load for demagnetization and eccentricity combination in per unit

Demagnetization	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ecc-01	No	0.94	0.59	0.72
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.43	0.47	0.45
		LDA	1	1	1
		PCA	1	1	1

50-20-1	Ecc-01	No	0.64	0.74	0.68
		LDA	0.94	1	0.97
		PCA	1	1	1
	Ecc-02	No	0.85	0.76	0.8
		LDA	1	1	1
		PCA	0.96	1	0.98
50-20-5	Ecc-01	No	0.71	0.8	0.75
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.57	0.55	0.56
		LDA	1	1	1
		PCA	1	1	1
50-5-5	Ecc-01	No	0.45	0.36	0.4
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.5	0.77	0.61
		LDA	0.96	1	0.98
		PCA	1	1	1

Table A-14: The important index of VAE using the sound sensor CH2 in rated speed and rated load for bearing and eccentricity combination in per unit

Bearing	Eccentricity	Reducing dimensionality	Precision	Recall	F1-Score
Ball-101	Ecc-01	No	0.66	0.86	0.75
		LDA	0.97	0.97	0.97
		PCA	0.9	1	0.95
	Ecc-02	No	0.36	0.45	0.4
		LDA	1	1	1
		PCA	1	1	1
Ball-2015	Ecc-01	No	0.64	0.45	0.53
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.48	0.74	0.58
		LDA	1	1	1
		PCA	1	1	1
Inner-101	Ecc-01	No	0.53	0.52	0.52
		LDA	1	1	1

		PCA	1	0.88	0.94
	Ecc-02	No	0.51	0.7	0.59
		LDA	0.97	1	0.98
		PCA	0.97	1	0.99
Outer-101	Ecc-01	No	0.49	0.74	0.59
		LDA	1	1	1
		PCA	1	1	1
	Ecc-02	No	0.55	0.33	0.41
		LDA	1	0.97	0.98
		PCA	1	1	1

Table A-15: The important index of VAE using the sound sensor CH2 in rated speed and rated load for demagnetization and bearing combination in per unit

Demagnetization	Bearing	Reducing dimensionality	Precision	Recall	F1-Score
10-20-5	Ball-101	No	0.6	0.2	0.3
		LDA	1	0.9	0.95
		PCA	1	1	1
	Ball-2015	No	0.66	0.7	0.68
		LDA	0.91	0.94	0.92
		PCA	0.97	1	0.98
	Inner-101	No	0.88	1	0.94
		LDA	0.97	0.97	0.97
		PCA	0.97	1	0.98
	Outer-101	No	0.51	0.83	0.63
		LDA	0.95	0.95	0.95
		PCA	0.96	1	0.98
50-20-1	Ball-101	No	0.64	0.68	0.66
		LDA	1	1	1
		PCA	1	1	1
	Ball-2015	No	0.58	0.37	0.45
		LDA	1	1	1
		PCA	1	0.97	0.99
	Inner-101	No	0.4	0.11	0.18
		LDA	0.94	1	0.97
		PCA	1	1	1

	Outer-101	No	1	0.47	0.64
		LDA	0.93	0.93	0.93
		PCA	1	0.96	0.98
50-20-5	Ball-101	No	0.78	0.67	0.72
		LDA	0.94	1	0.97
		PCA	1	1	1
	Ball-2015	No	0.63	0.61	0.62
		LDA	1	1	1
		PCA	1	1	1
	Inner-101	No	0.82	0.69	0.75
		LDA	0	0	0
		PCA	1	1	1
	Outer-101	No	0.52	0.64	0.57
		LDA	0.96	0.96	0.96
		PCA	1	0.97	0.98
50-5-5	Ball-101	No	0.66	0.88	0.75
		LDA	0.96	0.9	0.93
		PCA	1	1	1
	Ball-2015	No	0.84	0.96	0.9
		LDA	0.97	1	0.98
		PCA	1	1	1
	Inner-101	No	0.86	0.76	0.81
		LDA	1	0.88	0.93
		PCA	1	1	1
	Outer-101	No	0.65	0.71	0.68
		LDA	0.96	0.92	0.94
		PCA	1	0.94	0.97
-	-	No	0.43	0.39	0.41
		LDA	0.96	0.9	0.93
		PCA	1	1	1

Table A-16: The important index of VAE using the current sensor in rated speed and rated load for demagnetization, bearing and eccentricity combination in per unit

Demagnetization	Bearing	Eccentricity	reduce	Precision	Recall	F1-Score
10205	Ball101	Ecc-01	NO	0.83	0.83	0.83

			LDA	0.88	1	0.93
			PCA	1	1	1
		Ecc-02	NO	0.83	1	0.91
			LDA	0.94	0.97	0.95
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.48	0.48	0.48
			LDA	0.9	0.9	0.9
			PCA	1	1	1
		Ecc-02	NO	0.88	0.54	0.67
			LDA	0.86	0.81	0.83
			PCA	0.65	0.58	0.61
	Inner 101	Ecc-01	NO	0.82	0.78	0.8
			LDA	0.93	0.84	0.89
			PCA	1	1	1
		Ecc-02	NO	0.96	0.74	0.84
			LDA	1	1	1
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.61	0.85	0.71
			LDA	0.9	0.97	0.93
			PCA	1	1	1
		Ecc-02	NO	0.79	0.51	0.62
			LDA	0.78	0.75	0.77
			PCA	1	1	1
50201	Ball101	Ecc-01	NO	1	1	1
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0.77	0.9	0.83
			LDA	0.92	0.94	0.93
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.94	0.91	0.92
			LDA	1	0.97	0.99
			PCA	1	1	1
		Ecc-02	NO	1	1	1
			LDA	1	1	1

			PCA	1	1	1
	Inner 101	Ecc-01	NO	1	1	1
			LDA	0.94	1	0.97
			PCA	1	1	1
		Ecc-02	NO	1	0.97	0.98
			LDA	0.93	1	0.97
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.96	1	0.98
			LDA	0.97	0.97	0.97
			PCA	1	1	1
		Ecc-02	NO	1	1	1
			LDA	1	1	1
			PCA	1	1	1
50205	Ball101	Ecc-01	NO	1	1	1
			LDA	0.88	1	0.93
			PCA	1	1	1
		Ecc-02	NO	0.97	1	0.99
			LDA	0.97	1	0.99
			PCA	1	1	1
	Ball2015	Ecc-01	NO	1	0.97	0.98
			LDA	0.97	1	0.98
			PCA	1	1	1
		Ecc-02	NO	1	1	1
			LDA	0.96	0.81	0.88
			PCA	1	1	1
	Inner 101	Ecc-01	NO	1	1	1
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	1	1	1
			LDA	0.86	0.9	0.88
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.9	0.76	0.83
			LDA	1	1	1
			PCA	1	1	1

		Ecc-02	NO	0.9	0.93	0.92
			LDA	0.94	0.97	0.96
			PCA	1	1	1
5055	Ball101	Ecc-01	NO	0.84	0.84	0.84
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0.89	0.97	0.93
			LDA	0.84	0.96	0.9
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.82	0.53	0.64
			LDA	0.95	0.76	0.84
			PCA	1	1	1
		Ecc-02	NO	0.55	0.68	0.61
			LDA	0.88	0.91	0.89
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0.76	0.81	0.79
			LDA	0.89	0.94	0.91
			PCA	1	1	1
		Ecc-02	NO	0.91	0.97	0.94
			LDA	1	0.97	0.98
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.59	0.83	0.69
			LDA	0.91	0.97	0.94
			PCA	1	1	1
		Ecc-02	NO	0.59	0.87	0.7
			LDA	0.92	0.73	0.81
			PCA	0.53	0.6	0.56
-	-	-	NO	0.83	0.83	0.83
			LDA	0.88	1	0.93
			PCA	1	1	1

Table A-17: The important index of VAE using the horizontal Vibration sensor in rated speed and rated load for demagnetization, bearing and eccentricity combination in per unit

Demagnetization	Bearing	Eccentricity	reduce	Precision	Recall	F1-Score
10205	Ball101	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	0.96	0.92	0.94
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	0.97	0.97	0.97
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0	0	0
			LDA	0.97	0.97	0.97
			PCA	0.97	1	0.99
		Ecc-02	NO	0	0	0
			LDA	0.91	0.91	0.91
			PCA	1	1	1
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.93	1	0.96
		Ecc-02	NO	0	0	0
			LDA	1	0.94	0.97
			PCA	1	1	1
50201	Ball101	Ecc-01	NO	0	0	0
			LDA	1	0.89	0.94
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0

			LDA	1	0.97	0.98
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	0.97	0.99
			PCA	0.85	0.97	0.91
	Inner 101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.92	1	0.96
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	0.97	1	0.98
50205	Ball101	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	0.97	1	0.99
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0	0	0
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0	0	0
			LDA	0.96	0.9	0.93
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	0.83	1	0.91

			PCA	1	1	1
	Outer101	Ecc-01	NO	0	0	0
			LDA	0.86	0.96	0.91
			PCA	0.97	1	0.98
		Ecc-02	NO	0	0	0
			LDA	0.97	0.94	0.95
			PCA	1	1	1
5055	Ball101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	0.92	0.92	0.92
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.97	1	0.99
		Ecc-02	NO	0	0	0
			LDA	1	0.96	0.98
			PCA	0.96	1	0.98
	Inner 101	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	0.97	1	0.99
		Ecc-02	NO	0.02	1	0.04
			LDA	1	0.95	0.97
			PCA	0.67	1	0.8
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	0.92	1	0.96
-	-	-	NO	0	0	0

			LDA	1	1	1
			PCA	1	1	1

Table A-18: The important index of VAE using the vertical vibration sensor in rated speed and rated load for demagnetization, bearing and eccentricity combination in per unit

Demagnetization	Bearing	Eccentricity	reduce	Precision	Recall	F1-Score
10205	Ball101	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	1	1	0.98
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	0.96	0.96	0.96
			PCA	1	0.96	0.96
		Ecc-02	NO	0	0	0
			LDA	0.89	0.8	0.84
			PCA	0.61	0.8	0.84
	Inner 101	Ecc-01	NO	0	0	0
			LDA	0.96	0.96	0.96
			PCA	0.83	0.96	0.96
		Ecc-02	NO	0	0	0
			LDA	0.9	0.93	0.92
			PCA	0.47	0.93	0.92
	Outer101	Ecc-01	NO	0	0	0
			LDA	0.97	0.97	0.97
			PCA	0.53	0.97	0.97
		Ecc-02	NO	0	0	0
			LDA	0.96	0.9	0.93
			PCA	0.9	0.9	0.93
50201	Ball101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	1	1

			PCA	0.97	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	1	1	0.98
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0.01	1	0.02
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	0.96	0.98
			PCA	1	0.96	0.98
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.97	1	1
		Ecc-02	NO	0	0	0
			LDA	0.97	1	0.99
			PCA	1	1	0.99
50205	Ball101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.94	1	1
		Ecc-02	NO	0	0	0
			LDA	1	0.97	0.99
			PCA	0.97	0.97	0.99
	Ball2015	Ecc-01	NO	0	0	0
			LDA	0.97	1	0.98
			PCA	0.97	1	0.98
		Ecc-02	NO	0	0	0
			LDA	0.9	0.96	0.93
			PCA	0.97	0.96	0.93
	Inner 101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.97	1	1

		Ecc-02	NO	0	0	0
			LDA	1	0.96	0.98
			PCA	0.86	0.96	0.98
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	0.96	1	1
		Ecc-02	NO	0	0	0
			LDA	1	0.97	0.99
			PCA	1	0.97	0.99
5055	Ball101	Ecc-01	NO	0	0	0
			LDA	0.96	1	0.98
			PCA	0.9	1	0.98
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0	0	0
			LDA	0.96	1	0.98
			PCA	1	1	0.98
		Ecc-02	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
	Outer101	Ecc-01	NO	0	0	0
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0	0	0
			LDA	0.92	1	0.96
			PCA	1	1	0.96

-	-	-	NO	0	0	0
			LDA	1	0.93	0.96
			PCA	0.92	0.93	0.96

Table A-19: The important index of VAE using the sound sensor CH1 in rated speed and rated load for demagnetization, bearing and eccentricity combination in per unit

Demagnetization	Bearing	Eccentricity	reduce	Precision	Recall	F1-Score
10205	Ball101	Ecc-01	NO	0.42	0.23	0.3
			LDA	0.96	1	0.98
			PCA	0.96	1	0.98
		Ecc-02	NO	0.55	0.79	0.65
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.67	0.37	0.47
			LDA	0.85	1	0.92
			PCA	0.95	1	0.97
		Ecc-02	NO	0.69	0.83	0.76
			LDA	0.97	1	0.98
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0.83	0.73	0.78
			LDA	0.97	0.91	0.94
			PCA	1	0.93	0.96
		Ecc-02	NO	0.64	0.62	0.63
			LDA	0.93	1	0.96
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.75	0.7	0.72
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0.71	0.86	0.77
			LDA	1	0.9	0.95
			PCA	1	1	1
50201	Ball101	Ecc-01	NO	0.56	0.12	0.2
			LDA	0.77	0.9	0.83
			PCA	0.86	1	0.93
		Ecc-02	NO	0.44	0.43	0.44

			LDA	0.94	0.97	0.95
			PCA	0.96	0.85	0.9
	Ball2015	Ecc-01	NO	0.47	0.78	0.59
			LDA	0.9	0.79	0.84
			PCA	1	0.9	0.95
		Ecc-02	NO	0.61	0.44	0.51
			LDA	0.85	0.92	0.88
			PCA	0.9	1	0.95
	Inner 101	Ecc-01	NO	1	0.77	0.87
			LDA	0.97	1	0.99
			PCA	1	1	1
		Ecc-02	NO	0.75	0.51	0.61
			LDA	0.94	0.94	0.94
			PCA	0.94	0.97	0.95
	Outer101	Ecc-01	NO	0.65	0.9	0.75
			LDA	0.94	0.97	0.96
			PCA	0.96	1	0.98
		Ecc-02	NO	0.47	0.23	0.3
			LDA	0.96	1	0.98
			PCA	0.89	1	0.94
50205	Ball101	Ecc-01	NO	0.43	0.26	0.33
			LDA	1	0.97	0.98
			PCA	1	0.96	0.98
		Ecc-02	NO	0.46	0.79	0.58
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.64	0.77	0.7
			LDA	0.97	1	0.98
			PCA	1	1	1
		Ecc-02	NO	0.86	0.9	0.88
			LDA	0.95	0.56	0.7
			PCA	1	0.97	0.99
	Inner 101	Ecc-01	NO	0.61	0.87	0.71
			LDA	0.85	1	0.92

			PCA	1	0.85	0.92
		Ecc-02	NO	0.68	0.76	0.72
			LDA	0.86	1	0.93
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.61	0.52	0.56
			LDA	0.86	1	0.93
			PCA	1	1	1
		Ecc-02	NO	0.18	0.3	0.22
			LDA	1	0.9	0.95
			PCA	0.88	0.88	0.88
5055	Ball101	Ecc-01	NO	0.88	0.6	0.71
			LDA	1	0.96	0.98
			PCA	1	0.96	0.98
		Ecc-02	NO	0.48	0.42	0.45
			LDA	1	0.96	0.98
			PCA	0.97	0.91	0.94
	Ball2015	Ecc-01	NO	0.35	0.44	0.39
			LDA	1	0.94	0.97
			PCA	0.96	1	0.98
		Ecc-02	NO	1	1	1
			LDA	0.97	0.94	0.96
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0.94	1	0.97
			LDA	0.96	1	0.98
			PCA	1	1	1
		Ecc-02	NO	0.71	0.96	0.81
			LDA	0.94	1	0.97
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.24	0.32	0.27
			LDA	1	0.88	0.94
			PCA	0.96	1	0.98
		Ecc-02	NO	0.5	0.54	0.52
			LDA	1	1	1

			PCA	1	1	1
-	-	-	NO	0.77	0.88	0.82
			LDA	0.88	0.86	0.87
			PCA	1	1	1

Table A-20: The important index of VAE using the sound sensor CH2 in rated speed and rated load for demagnetization, bearing and eccentricity combination in per unit

Demagnetization	Bearing	Eccentricity	reduce	Precision	Recall	F1-Score
10205	Ball101	Ecc-01	NO	0.45	0.33	0.38
			LDA	1	1	1
			PCA	0.95	1	0.97
		Ecc-02	NO	0.59	0.77	0.67
			LDA	1	1	1
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.52	0.77	0.62
			LDA	1	0.96	0.98
			PCA	1	1	1
		Ecc-02	NO	0.79	0.7	0.74
			LDA	0.96	1	0.98
			PCA	1	1	1
	Inner 101	Ecc-01	NO	0.78	0.81	0.79
			LDA	0.89	0.97	0.93
			PCA	0.96	0.92	0.94
		Ecc-02	NO	0.94	0.52	0.67
			LDA	0.96	1	0.98
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.57	0.74	0.65
			LDA	1	0.97	0.98
			PCA	1	1	1
		Ecc-02	NO	0.84	0.73	0.78
			LDA	1	0.98	0.99
			PCA	1	1	1
50201	Ball101	Ecc-01	NO	0.52	0.34	0.41
			LDA	0.82	0.88	0.85
			PCA	1	1	1

		Ecc-02	NO	0.46	0.5	0.48
			LDA	1	0.94	0.97
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.64	0.69	0.67
			LDA	0.9	0.93	0.91
			PCA	0.97	1	0.98
		Ecc-02	NO	0.47	0.38	0.42
			LDA	0.8	1	0.89
			PCA	0.94	1	0.97
	Inner 101	Ecc-01	NO	0.92	0.96	0.94
			LDA	0.9	0.9	0.9
			PCA	1	1	1
		Ecc-02	NO	0.55	0.7	0.62
			LDA	0.9	0.96	0.93
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.78	0.89	0.83
			LDA	1	0.93	0.96
			PCA	1	1	1
		Ecc-02	NO	0.64	0.25	0.36
			LDA	0.97	0.97	0.97
			PCA	1	1	1
50205	Ball101	Ecc-01	NO	0.33	0.26	0.29
			LDA	1	0.9	0.95
			PCA	0.97	0.91	0.94
		Ecc-02	NO	0.59	0.59	0.59
			LDA	0.97	0.94	0.95
			PCA	1	1	1
	Ball2015	Ecc-01	NO	0.57	0.85	0.68
			LDA	0.91	0.97	0.94
			PCA	1	1	1
		Ecc-02	NO	0.92	0.92	0.92
			LDA	0.88	0.66	0.75
			PCA	1	0.97	0.99
	Inner 101	Ecc-01	NO	0.44	0.58	0.5

			LDA	0.96	0.96	0.96
			PCA	1	1	1
		Ecc-02	NO	0.7	0.68	0.69
			LDA	0.91	1	0.95
			PCA	0.97	1	0.98
	Outer101	Ecc-01	NO	0.62	0.75	0.68
			LDA	0.93	1	0.96
			PCA	1	1	1
		Ecc-02	NO	0.38	0.16	0.22
			LDA	1	1	1
			PCA	0.81	0.81	0.81
5055	Ball101	Ecc-01	NO	0.84	0.59	0.7
			LDA	0.84	0.95	0.89
			PCA	1	1	1
		Ecc-02	NO	0.43	0.33	0.38
			LDA	1	1	1
			PCA	0.89	0.97	0.93
	Ball2015	Ecc-01	NO	0.48	0.39	0.43
			LDA	1	0.97	0.99
			PCA	1	0.97	0.98
		Ecc-02	NO	1	1	1
			LDA	1	0.94	0.97
			PCA	1	1	1
	Inner 101	Ecc-01	NO	1	1	1
			LDA	1	1	1
			PCA	1	1	1
		Ecc-02	NO	0.74	0.9	0.81
			LDA	0.96	1	0.98
			PCA	1	1	1
	Outer101	Ecc-01	NO	0.42	0.45	0.43
			LDA	1	0.97	0.98
			PCA	1	0.89	0.94
		Ecc-02	NO	0.46	0.77	0.58

			LDA	0.97	1	0.99
			PCA	1	1	1
			No	0.84	0.88	0.86
-	-	-	LDA	0.87	0.76	0.81
			PCA	1	1	1

8 REFERENCES

[1] J. F. T. E. TRANSITION, "The history behind Germany's nuclear phase-out," 2021.

[2] G. C. Stone, E. A. Boulter, I. Culbert, and H. Dhirani, *Electrical insulation for rotating machines: design, evaluation, aging, testing, and repair* vol. 21: John Wiley & Sons, 2004.

[3] J. Faiz and E. Mazaheri-Tehrani, "Demagnetization modeling and fault diagnosing techniques in permanent magnet machines under stationary and nonstationary conditions: An overview," *IEEE Transactions on Industry Applications,* vol. 53, pp. 2772-2785, 2016.

[4] N. A. Spaldin, *Magnetic materials: fundamentals and applications*: Cambridge university press, 2010.

[5] P. C. Sen, *Principles of electric machines and power electronics*: John Wiley & Sons, 2007.

[6] J.-R. R. Ruiz, J. A. Rosero, A. G. Espinosa, and L. Romeral, "Detection of demagnetization faults in permanent-magnet synchronous motors under nonstationary conditions," *IEEE Transactions on Magnetics,* vol. 45, pp. 2961-2969, 2009.

[7] Available:https://dev.to/trekhleb/playing-with-discrete-fourier-transform-algorithm-in-javascript-53n5]

[8] D. Brown, B. Smith, B. Ma, and P. Campbell, "The dependence of magnetic properties and hot workability of rare earth-iron-boride magnets upon composition," *IEEE Transactions on Magnetics,* vol. 40, pp. 2895-2897, 2004.

[9] Y. Da, X. Shi, and M. Krishnamurthy, "A new approach to fault diagnostics for permanent magnet synchronous machines using electromagnetic signature analysis," *IEEE Transactions on Power Electronics,* vol. 28, pp. 4104-4112, 2013.

[10] J.-C. Urresty, J.-R. Riba, and L. Romeral, "A back-emf based method to detect magnet failures in PMSMs," *IEEE Transactions on Magnetics,* vol. 49, pp. 591-598, 2013.

[11] P. Quseiri Darbandeh, M. Ardebili, and M. Aliyari Shoorehdeli, "Demagnetized Magnet Diagnosis in a Double-Layer Winding Permanent Magnet Synchronous Machine using Wavelet Packet Transform," *Computational Intelligence in Electrical Engineering,* 2021.

[12] X. Xiao, C. Chen, and M. Zhang, "Dynamic permanent magnet flux estimation of permanent magnet synchronous machines," *IEEE Transactions on Applied Superconductivity,* vol. 20, pp. 1085-1088, 2010.

[13] J. Farooq, A. Djerdir, and A. Miraoui, "Analytical modeling approach to detect magnet defects in permanent-magnet brushless motors," *IEEE Transactions on Magnetics,* vol. 44, pp. 4599-4604, 2008.

[14] S. Moon, J. Lee, H. Jeong, and S. W. Kim, "Demagnetization fault diagnosis of a PMSM based on structure analysis of motor inductance," *IEEE Transactions on Industrial Electronics,* vol. 63, pp. 3795-3803, 2016.

[15] J. De Bisschop, P. Sergeant, A. Hemeida, H. Vansompel, and L. Dupré, "Analytical model for combined study of magnet demagnetization and eccentricity defects in axial flux permanent magnet synchronous machines," *IEEE Transactions on Magnetics,* vol. 53, pp. 1-12, 2017.

[16] A. Sarikhani and O. A. Mohammed, "Inter-turn fault detection in PM synchronous machines by physics-based back electromotive force estimation," *IEEE Transactions on Industrial Electronics*, vol. 60, pp. 3472-3484, 2012.

[17] J. De Bisschop, A. Abdallh, P. Sergeant, and L. Dupré, "Identification of demagnetization faults in axial flux permanent magnet synchronous machines using an inverse problem coupled with an analytical model," *IEEE Transactions on Magnetics*, vol. 50, pp. 1-4, 2014.

[18] T. Ishikawa, Y. Seki, and N. Kurita, "Analysis for fault detection of vector-controlled permanent magnet synchronous motor with permanent magnet defect," *IEEE transactions on magnetics*, vol. 49, pp. 2331-2334, 2013.

[19] J. Farooq, S. Srairi, A. Djerdir, and A. Miraoui, "Use of permeance network method in the demagnetization phenomenon modeling in a permanent magnet motor," *IEEE Transactions on Magnetics*, vol. 42, pp. 1295-1298, 2006.

[20] T. Raminosoa, J. Farooq, A. Djerdir, and A. Miraoui, "Reluctance network modelling of surface permanent magnet motor considering iron nonlinearities," *Energy Conversion and Management*, vol. 50, pp. 1356-1361, 2009.

[21] D. Torregrossa, A. Khoobroo, and B. Fahimi, "Prediction of acoustic noise and torque pulsation in PM synchronous machines with static eccentricity and partial demagnetization using field reconstruction method," *IEEE Transactions on Industrial Electronics*, vol. 59, pp. 934-944, 2012.

[22] S. Chakraborty, E. Keller, A. Ray, and J. Mayer, "Detection and estimation of demagnetization faults in permanent magnet synchronous motors," *Electric Power Systems Research*, vol. 96, pp. 225-236, 2013.

[23] J.-C. Urresty, J.-R. Riba, M. Delgado, and L. Romeral, "Detection of demagnetization faults in surface-mounted permanent magnet synchronous motors by means of the zero-sequence voltage component," *IEEE transactions on Energy conversion*, vol. 27, pp. 42-51, 2011.

[24] M. Zafarani, T. Goktas, and B. Akin, "A comprehensive magnet defect fault analysis of permanent-magnet synchronous motors," *IEEE Transactions on Industry Applications*, vol. 52, pp. 1331-1339, 2015.

[25] M. Zafarani, T. Goktas, and B. Akin, "A comprehensive analysis of magnet defect faults in permanent magnet synchronous motors," in *2015 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2015, pp. 2779-2783.

[26] J.-C. Urresty, J.-R. Riba, and L. Romeral, "Influence of the stator windings configuration in the currents and zero-sequence voltage harmonics in permanent magnet synchronous motors with demagnetization faults," *IEEE Transactions on Magnetics*, vol. 49, pp. 4885-4893, 2013.

[27] J. Rosero, J. Cusido, A. Garcia, J. Ortega, and L. Romeral, "Study on the permanent magnet demagnetization fault in permanent magnet synchronous machines," in *IECON 2006-32nd Annual Conference on IEEE Industrial Electronics*, 2006, pp. 879-884.

[28] M. Heydarzadeh, M. Zafarani, M. Nourani, and B. Akin, "A Wavelet-Based Fault Diagnosis Approach for Permanent Magnet Synchronous Motors," *IEEE Transactions on Energy Conversion*, vol. 34, pp. 761-772, 2018.

[29] J. Faiz and E. Mazaheri-Tehrani, "Demagnetization modeling and fault diagnosing techniques in permanent magnet machines under stationary and nonstationary conditions: An overview," *IEEE Transactions on Industry Applications*, vol. 53, pp. 2772-2785, 2017.

[30] S. Rajagopalan, J. A. Restrepo, J. M. Aller, T. G. Habetler, and R. G. Harley, "Nonstationary motor fault detection using recent quadratic time–frequency

representations," *IEEE Transactions on Industry Applications,* vol. 44, pp. 735-744, 2008.

[31] A. G. Espinosa, J. A. Rosero, J. Cusido, L. Romeral, and J. A. Ortega, "Fault detection by means of Hilbert–Huang transform of the stator current in a PMSM with demagnetization," *IEEE Transactions on Energy Conversion,* vol. 25, pp. 312-318, 2010.

[32] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical systems and signal processing,* vol. 35, pp. 108-126, 2013.

[33] M. D. Prieto, A. G. Espinosa, J.-R. R. Ruiz, J. C. Urresty, and J. A. Ortega, "Feature extraction of demagnetization faults in permanent-magnet synchronous motors based on box-counting fractal dimension," *IEEE Transactions on Industrial Electronics,* vol. 58, pp. 1594-1605, 2010.

[34] C. Wang, M. D. Prieto, L. Romeral, Z. Chen, F. Blaabjerg, and X. Liu, "Detection of partial demagnetization fault in PMSMs operating under nonstationary conditions," *IEEE Transactions on Magnetics,* vol. 52, pp. 1-4, 2016.

[35] S. Moon, H. Jeong, H. Lee, and S. W. Kim, "Detection and classification of demagnetization and interturn short faults of IPMSMs," *IEEE Transactions on Industrial Electronics,* vol. 64, pp. 9433-9441, 2017.

[36] I.-H. Kao, W.-J. Wang, Y.-H. Lai, and J.-W. Perng, "Analysis of permanent magnet synchronous motor fault diagnosis based on learning," *IEEE Transactions on Instrumentation and Measurement,* vol. 68, pp. 310-324, 2018.

[37] L.-K. Chang, S.-H. Wang, and M.-C. Tsai, "Demagnetization Fault Diagnosis of a PMSM Using Auto-Encoder and K-Means Clustering," *Energies,* vol. 13, p. 4467, 2020.

[38] L. Zhan, X. Xu, X. Qiao, F. Qian, and Q. Luo, "Fault Feature Extraction Method of a Permanent Magnet Synchronous Motor Based on VAE-WGAN," *Processes,* vol. 10, p. 200, 2022.

[39] J. Liu and Z. Zhu, "Improved sensorless control of permanent-magnet synchronous machine based on third-harmonic back EMF," *IEEE Transactions on Industry Applications,* vol. 50, pp. 1861-1870, 2013.

[40] K.-W. Lee, J. Hong, S. B. Lee, and S. Lee, "Quality assurance testing for magnetization quality assessment of BLDC motors used in compressors," *IEEE Transactions on Industry Applications,* vol. 46, pp. 2452-2458, 2010.

[41] R. Z. Haddad, C. A. Lopez, S. N. Foster, and E. G. Strangas, "A voltage-based approach for fault detection and separation in permanent magnet synchronous machines," *IEEE Transactions on Industry Applications,* vol. 53, pp. 5305-5314, 2017.

[42] R. Z. Haddad and E. G. Strangas, "On the accuracy of fault detection and separation in permanent magnet synchronous machines using MCSA/MVSA and LDA," *IEEE Transactions on Energy Conversion,* vol. 31, pp. 924-934, 2016.

[43] Z. Yang, X. Shi, and M. Krishnamurthy, "Vibration monitoring of PM synchronous machine with partial demagnetization and inter-turn short circuit faults," in *2014 IEEE Transportation Electrification Conference and Expo (ITEC)*, 2014, pp. 1-6.

[44] T. Ishikawa and N. Igarashi, "Failure Diagnosis of Demagnetization in Interior Permanent Magnet Synchronous Motors Using Vibration Characteristics," *Applied Sciences,* vol. 9, p. 3111, 2019.

[45] Y. Huang, C.-H. Chen, and C.-J. Huang, "Motor fault detection and feature extraction using RNN-based variational autoencoder," *IEEE Access*, vol. 7, pp. 139086-139096, 2019.

[46] D. Reigosa, D. Fernández, Y. Park, A. B. Diez, S. B. Lee, and F. Briz, "Detection of demagnetization in permanent magnet synchronous machines using hall-effect sensors," *IEEE Transactions on Industry Applications*, vol. 54, pp. 3338-3349, 2018.

[47] A. Khoobroo and B. Fahimi, "A novel method for permanent magnet demagnetization fault detection and treatment in permanent magnet synchronous machines," in *2010 Twenty-Fifth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2010, pp. 2231-2237.

[48] C. Gao, Y. Nie, J. Si, Z. Fu, and H. Feng, "Mode Recognition and Fault Positioning of Permanent Magnet Demagnetization for PMSM," *Energies*, vol. 12, p. 1644, 2019.

[49] O. Zandi and J. Poshtan, "Fault Diagnosis of Brushless DC Motors Using Built-in Hall Sensors," *IEEE Sensors Journal*, 2019.

[50] J. Song, J. Zhao, F. Dong, J. Zhao, L. Xu, and Z. Yao, "A New Demagnetization Fault Recognition and Classification Method for DPMSLM," *IEEE Transactions on Industrial Informatics*, 2019.

[51] B. Kim, K. Lee, J. Yang, S. B. Lee, E. J. Wiedenbrug, and M. R. Shah, "Automated detection of rotor faults for inverter-fed induction machines under standstill conditions," *IEEE Transactions on Industry Applications*, vol. 47, pp. 55-64, 2010.

[52] R. Z. Haddad and E. G. Strangas, "Fault detection and classification in permanent magnet synchronous machines using Fast Fourier Transform and Linear Discriminant Analysis," in *Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), 2013 9th IEEE International Symposium on*, 2013, pp. 99-104.

[53] J. Hong, S. Park, D. Hyun, T.-j. Kang, S. B. Lee, C. Kral, *et al.*, "Detection and classification of rotor demagnetization and eccentricity faults for PM synchronous motors," *IEEE Transactions on Industry Applications*, vol. 48, pp. 923-932, 2012.

[54] B. M. Ebrahimi and J. Faiz, "Demagnetization fault diagnosis in surface mounted permanent magnet synchronous motors," *IEEE transactions on Magnetics*, vol. 49, pp. 1185-1192, 2012.

[55] M. Zhu, B. Yang, W. Hu, G. Feng, and N. C. Kar, "Vold–Kalman Filtering Order Tracking Based Rotor Demagnetization Detection in PMSM," *IEEE Transactions on Industry Applications*, vol. 55, pp. 5768-5778, 2019.

[56] Z. Ullah, S.-T. Lee, and J. Hur, "A Torque Angle based Fault Detection and Identification Technique for IPMSM," *IEEE Transactions on Industry Applications*, 2019.

[57] M. Zhu, W. Hu, and N. C. Kar, "Torque-ripple-based interior permanent-magnet synchronous machine rotor demagnetization fault detection and current regulation," *IEEE Transactions on Industry Applications*, vol. 53, pp. 2795-2804, 2016.

[58] T. M. Mitchell, "Machine learning," ed: McGraw-Hill, Inc., New York, NY, USA, 1997.

[59] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381-386, 2020.

[60] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering,* vol. 160, pp. 3-24, 2007.

[61] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern recognition,* vol. 40, pp. 2038-2048, 2007.

[62] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry,* vol. 27, p. 130, 2015.

[63] K. Koutroumbas and S. Theodoridis, *Pattern recognition*: Academic Press, 2008.

[64] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection," *IEEE/ACM transactions on computational biology and bioinformatics,* vol. 13, pp. 971-989, 2015.

[65] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science,* vol. 2, pp. 559-572, 1901.

[66] T. Cacoullos, *Discriminant analysis and applications*: Academic Press, 2014.

[67] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer,* vol. 29, pp. 31-44, 1996.

[68] W. McCullock and W. Pitts, "A logical calculus of ideas immanent in nervous activity. archive copy of 27 november 2007 on wayback machine," *Avtomaty [Automated Devices] Moscow, Inostr. Lit. publ,* pp. 363-384, 1956.

[69] H. Wang and B. Raj, "On the origin of deep learning," *arXiv preprint arXiv:1702.07800,* 2017.

[70] V. Kazak, "Unsupervised feature extraction with autoencoder: for the representation of parkinson´ s disease patients," 2019.

[71] S. Torres Fernandez, "Designing Variational Autoencoders for Image Retrieval," ed, 2018.

[72] S. Haykin, *Neural networks and learning machines, 3/E*: Pearson Education India, 2009.

[73] B. Widrow and M. Hoff, "Adaptive switching circuits, IRE WESCON Conv," *Rec., pt,* vol. 4, p. 96, 1960.

[74] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature,* vol. 323, pp. 533-536, 1986.

[75] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*: MIT press, 2016.

[76] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning,* vol. 12, pp. 307-392, 2019.

[77] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS),* 2018, pp. 1-2.

[78] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems,* vol. 31, 2018.

[79] G. Dajaku and D. Gerling, "Design of permanent magnet machines for hybrid vehicles," in *Congress Center Dresden, Germany,* 2007.

[80] P. C. Krause, O. Wasynczuk, S. D. Sudhoff, and S. D. Pekarek, *Analysis of electric machinery and drive systems* vol. 75: John Wiley & Sons, 2013.

[81] X. Wang and C. S. Suh, "Time-frequency based field oriented control of permanent magnet synchronous motors," *International Journal of Dynamics and Control,* vol. 6, pp. 663-678, 2018.