## Chapter 2

# Literature Survey

## 2.1 Introduction

The comprehensive literature review on dragline, FMECA, and fault analysis has been presented in this chapter. The introduction to fault analysis, basic terminologies used in fault analysis, and classification of fault analysis approaches are briefly discussed. The research work focuses on data-driven based fault analysis of dragline; therefore, literature on unsupervised learning based fault analysis, semi-supervised learning based fault analysis and supervised learning based fault analysis have also been explained. Finally, this chapter contains literature survey of data-driven based fault analysis approaches with specific emphasis on BN based fault analysis and ANN based fault analysis.

## 2.2 Literature on Dragline

Dragline is used for stripping overburden in large opencast coal mines. The history of development of dragline and its deployment in opencast coal mines, and research works carried out on draglines have been explained in sub-sections 2.2.1 and 2.2.2.

### 2.2.1 History of dragline development

John W. Page invented the first dragline in 1904 as a partner of the firm Page & Schnabel Contracting used for digging of Chicago Canal [59]. Followed by this, the Bucyrus Erie, an American surface and underground mining equipment company, introduced the first crawler mounted dragline in 1911 and used electricity as a power source in 1912 [59]. After this, the Page Engineering Company built the first crude walking dragline with mobility in 1923 [60]. The dragline mining was initially introduced in India in early 1960s, and the first walking dragline was commissioned at Kurashia in South-Eastern Coalfields Limited, Korba, Chhattisgarh, in 1961 [59]. As of 2020, 43 draglines deployed in India [61], out of which 23 draglines are working in the Northern Coalfields Limited (NCL) [62]. In India, Heavy Engineering Corporation (HEC) is progressively manufacturing W-2000 model walking draglines indigenously in collaboration with Ransomes & Rapier. As of 2020, HEC supplied 15 draglines to the Indian mining industry [63]. The coal India limited has now standardized the draglines in two sizes, which are 10/70 and 24/96 for their mines [59]. Their economic life has been assumed to be 27 years [64]. In India, the NCL is the only subsidiary company of CIL, where the entire coal production is mined by opencast mining method and about 40% of the large volume of excavation is done through the use of large walking draglines [59].

### 2.2.2 Research works conducted on dragline

Research works on draglines are conducted to improve their performance either by reducing the occurrence of unwanted failure or by using reliability-centered maintenance policy to optimize the preventive maintenance interval of the dragline components [9], [15],

[16]. Studies conducted on draglines also emphasized on the lifetime parameter and failure mechanism [17], age-replacement model [5], or risk of failure components [18]. The time counter algorithm is also used to optimize the inspection interval of the dragline components [6], [19]. For diagnosing faults, the expert system is developed considering 18 components of dragline [29].

As it is presented in Table 1.1, the research works are conducted on various components of dragline. The failure mechanism of the swing pinion shaft due to non-uniform surface hardening of the tooth, that initiates the crack in the reasonable point of the pinion to reduce the fatigue strength in the root of the teeth, is identified [20]. The finite element analysis is used to identify the most probable stress in the critical locations of the bucket under the static and dynamic load conditions to prevent the failure of the bucket of the dragline [22]. The dynamic response of the bucket of the dragline is analyzed to improve the rigging design, and coordinate the motion of hoist and drag motors using experiment study in the laboratory to develop a mathematical model [21]. The structural health monitoring of draglines is done using lamb waves for both the diffraction pattern and the reflected waves to detect the damage and size of the crack in a cluster of the draglines [26]. The comparative study of several existing design codes is carried out to predict of fatigue life of the dragline cluster [27]. The relationship between stress concentration factors and weld roots failure of four full-size tubular joints of dragline is developed through laboratory testing to measure the weld profile and weld root gaps using the silicon imprint technique and feeler gauges [24]. The three-dimensional (3D) model is developed to predict the fatigue life of the front-end structure and operation performance of the dragline using Lagrange's equations and finite element analysis [25]. The crack of the weld joint of the cluster of the boom is detected to take the necessary repairing action to keep them in service using residual stress analysis [4]. The investigation of the fracture mechanism of wires at the socket termination of dragline boom support strands is carried out to identify the fatigue crack using radiography and electromagnetic testing methods [23]. However, most of the studies on dragline are conducted based on experimental analysis or comparative analyses, which are very costly and time-consuming. Hence data based fault analysis approach using artificial intelligence methods can be an effective alternative to identify the critical components and to predict the occurrence of faults in the dragline system.

## 2.3 Literature Review on FMECA

The FMECA is a quantitative analysis of the criticality of individual failure components of the system, which was developed in 1940s by the US military [65]. The FMECA links the likelihood of occurrence of failure, and its impact on the production and performance of the system to estimate the RPN [66]. The FMECA analyzes the risk of failure components, which is measured by criticality to take the appropriate preventive maintenance action that can reduce the possibility of occurrence of failure [67]. The wide industrial application of FMECA can be observed in the areas of energy generation industries [65], [68], [69], wind turbines [70]–[72], hydraulic systems [73], sludge treatment industries [74], semiconductor industries [75], automobile and vehicle industries [76], [77], and in machine tools [78]. In mining industry, FMECA based reliability modeling is used to identify the risk of various subsystems of the load-haul-dumper (LHD) machine deployed in underground mining focusing on their root causes [79]. Moreover, the FMEA is used in the yacht system design to rank the failure modes depending on the risk priority number to take corrective actions [80]. In this thesis, FMECA is used to identify the failure modes of dragline and estimate the RPN of its components.

## 2.4 Introduction to Fault Analysis

Over the last few decades, there is a rising demand to improve the reliability, performance, maintainability and fault control in the HEMM. The fault analysis of HEMM reveals sufficient information about the occurrence of fault, its quick and accurate detection and identification; thus providing economic operation [81]–[83]. Three basic variables (e.g., cause, symptom, and fault) are used in fault analysis. Fault occurs before the system leads to failure, and it reduces the performance of the system. The sequence of occurrence of these variables is presented in Figure 2.1.

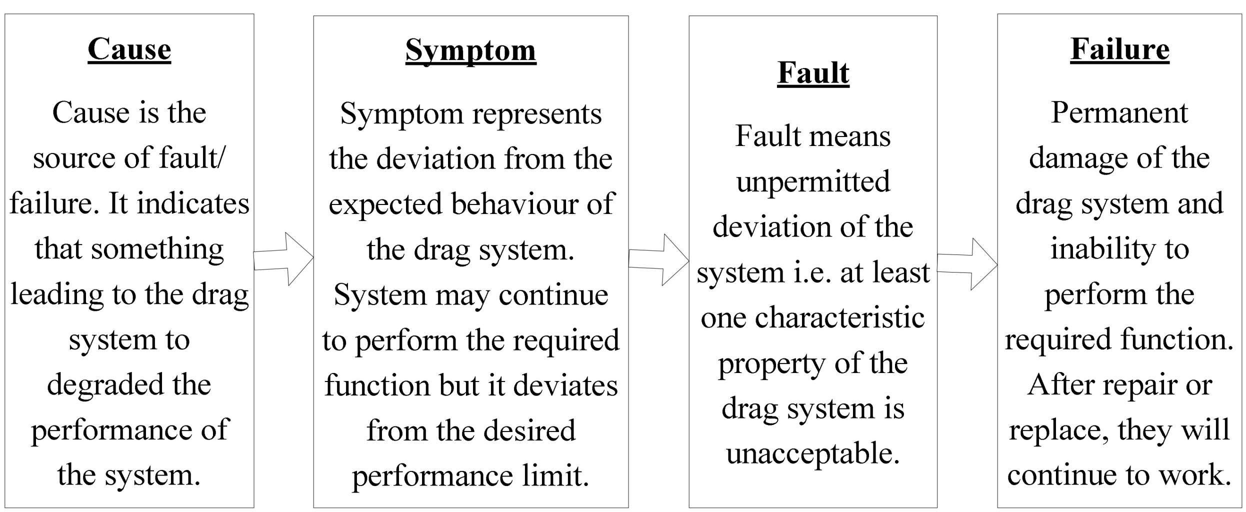


Figure 2.1 The sequence of occurrence of failure in the system

The presence of cause is the source of the fault and failure, and cause can be identified with the help of observed evidence (symptom). The occurrence symptom of the HEMM is usually detected through suitable sensors, visual inspection, or expert's opinion. The unmanaged causes can lead to a fault and subsequently can convert into a failure. In Figure 2.2, various stages of occurrence of failure are presented [84]. It also shows how the performance of the system degrades over time. The symptom zone, fault zone, and failure zones are also depicted to understand the process of failure occurrence. The maintenance should be taken with the potential failure to the functional failure (P-F) interval of the component to prevent the failure of the system. The P-F interval is the time interval between the point at which the initial symptom can be detected and the point at which functional failure occurs, which is shown in Figure 2.2 [13].

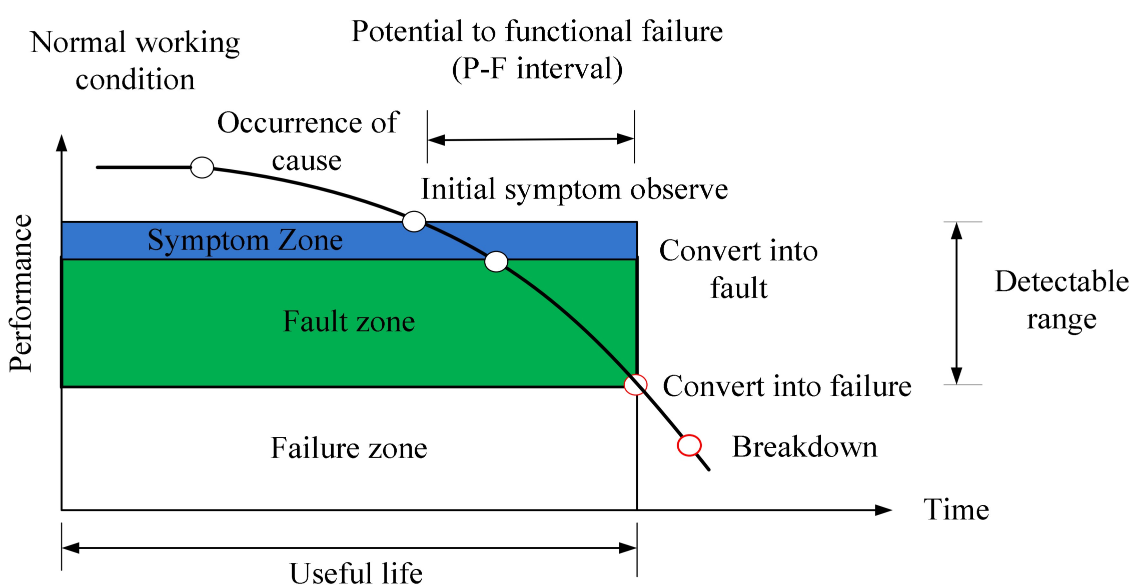


Figure 2.2 Graphical representation of the stages of occurrence of failure [84].

The P-F interval of the catastrophic fault is very less, hence urgent maintenance is required; on the other hand, the P-F interval of the degraded fault is large hence maintenance action can be performed during scheduled inspection interval or during the idle time to optimize the downtime of the system. The detectable range of fault is the time when the initial symptom is observed to degradation of the system to complete failure. In the detectable range, the system can perform the required function with unpermitted deviation, and its range is dependent on the nature, severity, and types of fault, which is shown in Figure 2.2. When the affected components are not isolated within the detectable range, it may damage the whole system connected in series [85]. These failures may cause unexpected breakdowns, downtime, high economic costs, and health and safety issues [86].

The fault analysis process consists of fault detection, fault isolation, fault identification, fault classification, based on the available evidence that helps to prepare a suitable maintenance policy. Fault detection is the process to detect malfunctions in real-time and check whether the system is running normally or not as surely as possible [81], [82], [87], [88]. The fault isolation process is used to finding the source of the fault, i.e. root causes and to start the appropriate preventive maintenance action to avoid subsequent failure [89]. The fault identification is the task to estimate the fault state that can best explain the behavior of the system to estimate the size and type or nature of the fault [58], [90]. The combined process of fault detection, isolation, and fault identification is known as fault diagnosis as shown in Figure 2.3 [88], [91]. The fault classification is a task to classify the fault based on type, nature, and effect on the system to prepare a suitable maintenance policy. The combined process of fault detection, isolation, fault identification, and fault classification is known as fault analysis (Figure 2.3).

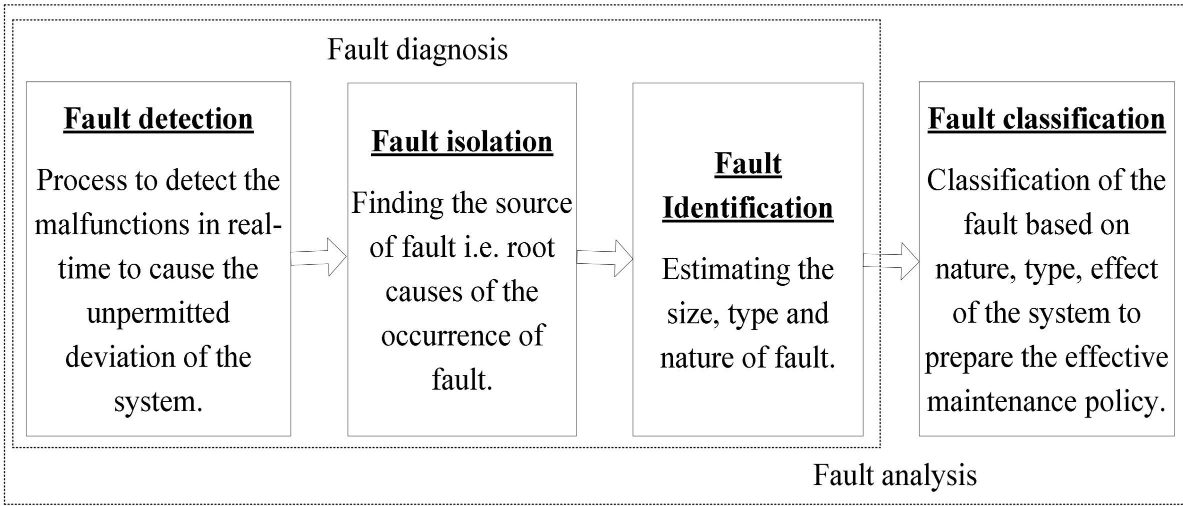


Figure 2.3 Process of fault analysis in the system

## 2.5 Classification of Fault Analysis Approaches

The fault analysis approaches can be divided into three categories: (i) rule-based fault analysis approach, (ii) model-based fault analysis approach, and (iii) data-driven fault analysis approach.

- The rule-based fault analysis approach is used when the system is simple to follow the specific rule such as if-then-else. The limitation of rule-based fault analysis approach is the system follows the defined rules, e.g., electrical switching system [92].
- The model-based fault analysis approach is used to compare the actual behaviour of the system with model behaviour and then to identify the fault. The limitation of model-based fault analysis is to know the exact behaviour of the

system in normal condition as well as in faulty condition, e.g., actuator, sensor, and signal analysis [93].

- The data-driven based fault analysis approach is used to predict the occurrence of a fault and to identify the responsible root causes of the system based on observed evidence through the collected historical and real-time monitoring fault data. The data-driven fault analysis also uses acquired experience to solve new problems; hence it is also known as case-based reasoning approach [94].

Dragline is a large, complicated, and complex system, where one fault can be linked to multiple symptoms or multiple causes, and one cause can be responsible for the occurrence of multiple symptoms or multiple faults. Hence rule-based and model-based fault analysis approaches are not applicable to diagnose the faults efficiently [95]. The data-driven based fault analysis is one of the most powerful approaches used to predict the occurrence of fault and failure of the system [96], [97]. Access to a large amount of data of dragline is usually generated through the combination of maintenance worksheet, visual inspection, sensor data, and experts' opinions. The process of data collection, also known as data acquisition, may contain information about the system alarms history, root causes, symptoms of fault and failure, maintenance action, and administrative structure of the management. After data acquisition, its processing (e.g., data cleaning, missing data treatment, removal of outlier, data transformation), and data categorization is done based on their threshold limit value [90], [98]. These data can be learnt using advanced artificial intelligence tools to extract meaningful information to construct the model to diagnose the fault [96], [99]. After that validation of the model significantly provides reasonable confidence of the diagnostic results of fault analysis methodology. Results of fault analysis can be used for defining the decision support system to decide the best maintenance policy and timely maintenance based on the available evidence, without degrading the performance, safety of the equipment that optimizes the maintenance cost of the system [13], [100].

## 2.6 Classification of Data-Driven Based Fault Analysis Approach

Based on the available information of the system, the data-driven based fault analysis approaches are categorized into three categories: unsupervised learning (USL) based fault analysis, semi-supervised learning (SSL) based fault analysis and supervised learning (SL) based fault analysis, as depicted in Figure 2.4 [101]–[104].

The data-driven based fault analysis is considered, and comprehensive literature survey has been carried out using available search engines using a combination of keywords, including fault detection, fault isolation, fault diagnosis, fault classification, fault identification, and fault analysis.

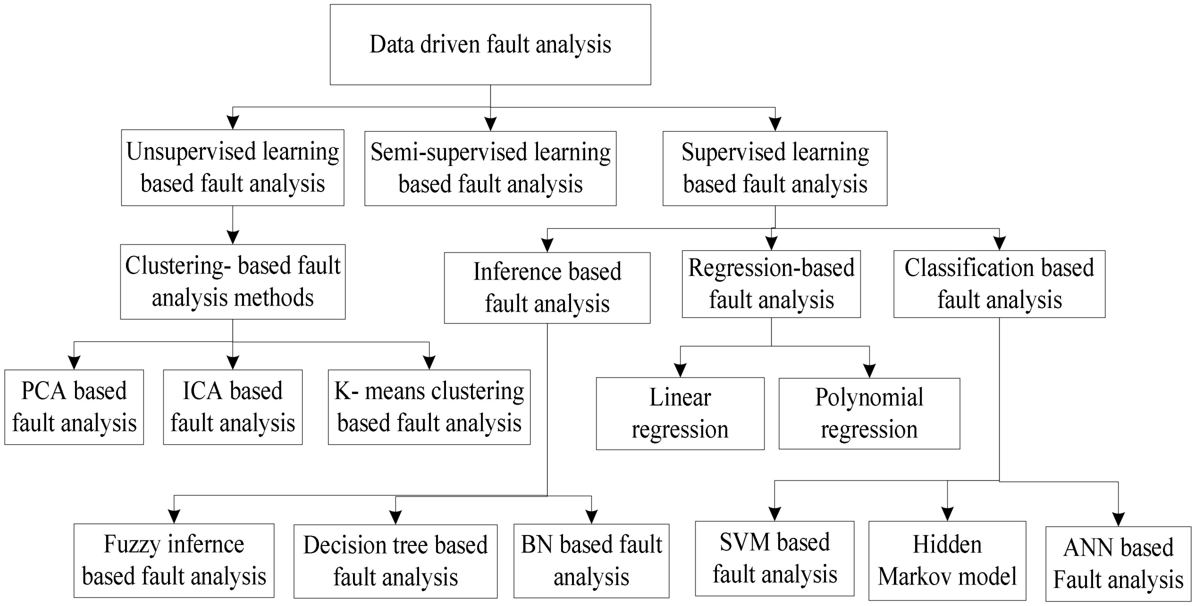


Figure 2.4 Classification of data-driven fault analysis approach for HEMM.

There are 176 research papers on fault analysis approaches, those are used in various industrial applications starting from the year 1976 to 2020. Out of the 176 studies, 44 studies used unsupervised learning clustering-based fault analysis that includes 21 studies emphasized on principal component analysis (PCA), 15 on independent component analysis (ICA) and eight studies on K-means clustering algorithm for fault analysis (see Figure 2.5). There are only seven literature on semi-supervised learning based fault analysis methodology (Figure 2.5). There are 125 literature on supervised learning based fault analysis, and they are divided into three categories: inference based fault analysis (52 studies), regression-based fault analysis (seven studies), and classification based fault analysis (66 studies) (Figure 2.5). More specifically, inference based fault analysis technique is dominated by BN model (number of studies = 31), followed by the fuzzy logic 14, and for fault tree (FT) model, it is 7 (Figure 2.5). In the classification based fault analysis technique, there are 20 studies on support vector machine (SVM), 11 on Hidden Markov model (HMM), and 35 studies used ANN models for fault analysis.

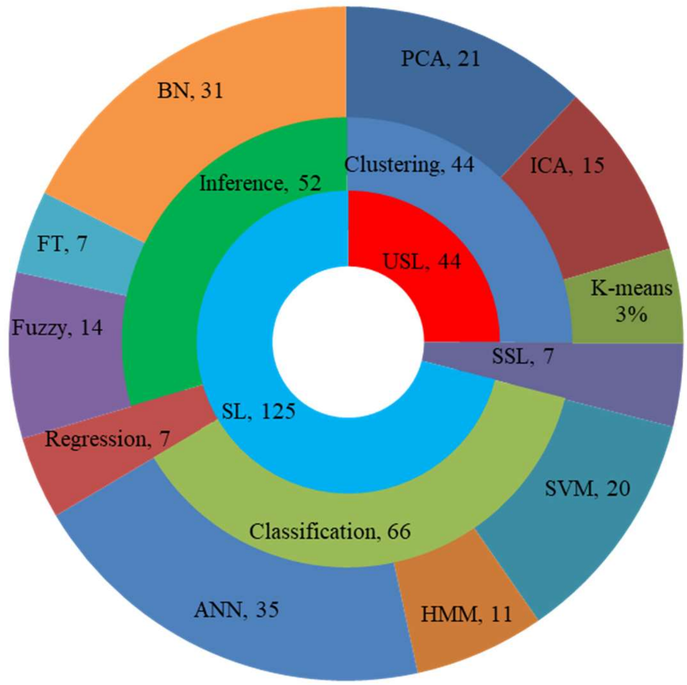


Figure 2.5 Classification of literature on data-driven based fault analysis

### 2.6.1 Unsupervised learning based fault analysis approach

The unsupervised learning based fault analysis approach is preferred when the unlabeled input data is collected from real-time monitoring, and the task of the system is not well defined [105]. The unlabeled collected data is mostly used in the clustering algorithm to identify the faults by dividing the collected data into several clusters according to their geometric distances and statistical characteristics between them [106]–[108]. The clustering-based fault analysis technique is mostly used in the exploratory data analysis to identify the intuition about the structure of the collected data using the PCA, ICA, and K-means clustering algorithm [109], [110], described in the following sub-sections.

### *2.6.1.1 Principal component analysis*

The PCA based fault analysis approach is used when the available process measurements are highly correlated, but only a small number of events (faults) produce unusual patterns [111], [112]. When the process data are highly correlated, the original process data represents the principal component subspace (normal condition), otherwise it represents the residual subspace (faulty condition) [113], [114]. The basic assumption in the use of PCA is all variables considered for analysis are normally distributed. The applications of PCA in the fault analysis found in the literature are extended to many industrial case studies, e.g., Tennessee Eastman process [115], air-conditioning system [116], steam generator of a nuclear power plant [117], wind turbine [118], hydraulic drive system [119], chiller sensor fault [49], gyroscopes [120], instrumental fault [121], vapour compression system [122], air quality monitoring network [123], and wastewater treatment plant [124], [125]. The limitations of PCA based fault analysis is that it needs to standardize the collected dataset before implementing PCA; otherwise, the principal component is not being to detect the fault [126]. The PCA is not suited for process monitoring based fault analysis of the system that displays non-stationary behaviour because they can break the correlation between the variables and produce excessive false alarms or missed faults detection [125].

### *Independent component analysis*

The ICA based fault analysis of the system differs from PCA. The PCA is applied when all variables follow normal distribution; whereas the ICA is preferred when all the variables of the system do not follow the normal distribution. Most of the complex systems consist of a large number of dependent and non-dependent variables, which drive a few independent variables and monitoring these variables using ICA can improve the performance of the system. The ICA is used to separate the statistically independent components from a set of available information [127], [128]. The ICA technique is also used as a pre-processing tool of SVM or factorial HMM [50], [129]. Various applications of the ICA based fault analysis are found in the literature, e.g., in turbine of the thermal power plants [52], machine fault [50], continuous-stirred-tank-reactor process [127], [130], wastewater treatment process [128], Tennessee Eastman process [126], [131]–[134], gyroscope system [135], and condenser system in the thermal power plants [51]. Though it poses certain advantages over the PCA, the limitations of ICA are that it slows down the capabilities of approximations and leads to slow convergence. The ICA algorithms are sensitive to the measurement of noise and magnitude, hence it is difficult to identify the fault in those systems [126].

### *K-means clustering*

The K-means clustering-based fault analysis is used to partition the collected dataset into a fixed number of clusters (K partitions) to assign data points in the cluster and minimizes the sum of the squared distance between the data points and the cluster's centroid. The advantage of the K-means clustering is its easy implementation and good performance for handling large dataset [136], [137]. The K-means clustering-based fault analysis technique is used in various industrial applications, such as the chiller water system [138], marine engine system [139], bearing fault [53], [140], and aluminium smelting process [54]. The limitation of K-means is that it can only be used to handle numerical data, and the results might be skewed if the data are not normalized. Selecting the number of clusters (K) is a difficult task when the prior knowledge of collected dataset is unknown. The basic assumption of K-means clustering-based fault analysis is to deal with spherical clusters and that each cluster has roughly equal numbers of observations [141].

### 2.6.2 Semi-supervised learning based fault analysis approach

The semi-supervised learning based fault analysis is preferred when the collection of fully labelled data is difficult to identify, or it is very costly [142]–[144] and some label data (training data) is used to improve the learning accuracy of the system [145]. Hence the semi-supervised learning is a combination of both supervised and unsupervised learning. Depending on the nature of the output parameter, the SSL based fault analysis approach is divided into two categories: semi-supervised regression and semi-supervised classification [145]. The semi-supervised learning based fault analysis is used in various industrial applications, such as the air handling unit [144], planetary gearbox [146], complex process [147], industrial process [142], bearing fault [148], nuclear power plants [149], and power transformer [150].

### 2.6.3 Supervised learning based fault analysis approach

The supervised learning based fault analysis is preferred when the fully labelled data is available (i.e. input and output parameter of the system are known) to analyze the hidden patterns to make relationships between them [35]. The supervised learning based fault analysis approach is divided into three groups: inference based fault analysis, regression-based fault analysis, and classification based fault analysis.

#### 2.6.3.1 Inference based fault analysis

Inference based fault analysis is used to draw a conclusion based on the observed set of evidence. Inference based fault analysis mostly uses three approaches: fuzzy inference, fault tree, and BN.

*Fuzzy inference*

The fuzzy set theory is mostly used to identify a fault from the real-time monitoring data to describe the behaviour of the equipment concerned in the current operating condition and to isolate the fault using some linguistic rules [151], [152]. The key advantage of fuzzy logic is to describe the system behaviour by simple 'if-then' relationship, and the operator can easily understand the faults and fault symptom relationship to take necessary action. The fuzzy inference based fault analysis is used in various industrial applications, e.g., rolling element of the bearings [153], cooling system of a diesel engine [154], fan coil [155], cooling-coil of a subsystem of the air-conditioning system [156], steam turbine [157], power distribution substations [38], direct current motor [151], evaporation station [152], induction machine [158], gas-turbine [159], Tennessee Eastman process [160], continuously-stirred tank heater [161], and continuously-stirred tank reactor [162]. The limitations of fuzzy inference based fault analysis is that the setting of exact fuzzy rules and defining membership functions is a complicated task to identify the fault. The validation and verification of the fuzzy rules required extensive testing that is not always possible, so it is not widely accepted.

### *Fault tree*

The fault tree, also known as decision tree, based fault analysis is a top-down, deductive fault analysis approach in which an undesired state of a system is analyzed. The decision tree does not require normalization and scaling of the data sample. The fault tree is used in the casual analysis of the system representing the interrelationships between a critical system event and its causes using Boolean logic [39]. It is also used to monitor and control the safety performance of the complex systems. The decision tree maps the relationships between faults, subsystems, and redundant safety design elements by creating a logic diagram of the overall system. The decision tree based fault analysis is used in various applications, such as in two-stroke marine diesel engine [163], nuclear power plants [164], longwall shearer [165], air handling unit [166], and belt conveyor system [167]. The limitation of fault tree analysis is that it is time taking, and calculation is more complicated as compared to other fault analysis approaches. The fault tree analysis is very sensitive when the small change in input data causes the large change in the structure of the decision tree causing instability of the model.

### *Bayesian Network*

In past few decades, data-driven based BN approaches have been extensively used in fault analysis applications to make the probabilistic graphical models to represent the causal relationships between multiple causes, multiple symptoms, and multiple faults [40], [168]–

[170]. The BN explains the structure of problems clearly with graph theory and problems to simplify the complexity of probabilistic reasoning [171], [172]. The BN can also work as a classifier based on Bayes' theorem and the conditional independence assumption [171], [173], [174]. The BN model is also used to predict the remaining useful life of the system [175]. The BN based fault analysis used in the various applications is presented in Table 2.1.

Table 2.1 BN based fault analysis approach of the various applications

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 1. | air-conditioning system | A two-layer BN model is used for the fault detection and diagnosis of variable air volume terminals of the air-conditioning system using a causal relationship between symptom and fault. | Xiao et al. [176] |
| 2. | Heat pump | Two-layer BN model used for fault diagnosis of ground-source heat pump using causal relationship between fault layer and fault symptom layer. | Cai et al. [177] |
| 3. | Rotating machinery | A three-layer BN constructed for the fault inference using experts' knowledge, machine faults and fault symptoms as well as machine running identify the root cause to parring the CBM policy. | Xu [7] |
| 4. | Centrifugal compressor | A three-layer BN model constructs to make a causal relationship between cause, symptom and fault using CPT to diagnose the fault and make the CBM policy. | Jun and Kim [13] |
| 5. | Centrifugal intelligent chiller | A three-layer BN model is used for chiller fault diagnosis to make a causal relationship between causal factors, faults, and fault symptoms, that help to prepare the CBM policy. | Zhao et al. [40] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 6. | Centrifugal chiller | Fault detection and diagnosis of chiller using BN classifier based on the false alarm rate using probabilistic boundary of BN model. | He et al. [171] |
| 7. | Centrifugal chiller | Fault diagnosis of centrifugal chiller using BN, merged distance rejection, and feature selection method. | Wang et al. [178] |
| 8. | Centrifugal chiller | Multi-source non-sensor information used in fault detection and diagnosis of chillers using merged distance rejection, and BN model. | Wang et al. [179] |
| 9. | Complex electronic systems | Fault identification of complex electronic system to distinguish the occurrence of faults as no fault, transient fault, intermittent fault, and permanent fault using dynamic BN model. | Cai et al. [180] |
| 10. | Subsea Christmas tree | Dynamic BN model is used to identify the faulty components and distinguish the fault types of subsea Christmas tree using failure rate and degradation characteristic of failure components. | Liu et al. [181] |
| 11. | Subsea production system | Real-time fault diagnosis model of subsea production system is developed using real-time sensor data, expert knowledge, operator experience for the faults and symptoms data using object oriented BN. | Cai et al. [182] |
| 12. | Tennessee Eastman Process | The HMM detects the abnormalities while the BN diagnoses the root causes of faults using prior process flow sheet knowledge and causal relationship among different variables of the system. | Don and Khan [183] |
| 13. | Tennessee Eastman Chemical process | Process monitoring approach is used in the fault diagnosis and root cause identification of Tennessee Eastman Chemical process using dynamic BN model. | Mori and Yu [184] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 14. | Offshore wind farms | Fault diagnosis of wind turbine components of offshore wind farms based on vibration, temperature, and oil particle using BN to prepare CBM policy. | Asgarpour and Sorensen [185] |
| 15. | Wind turbine bearing fault | Fault diagnosis and prognosis of bearing fault using Bayesian inference to detect the probability of fault to make the predictive maintenance policy and estimate the remaining useful life. | Wang et al. [186] |
| 16. | Assembly fixture fault | Fault detection and localization of assembly fixture faults using multiple sensor data to construct BN model and also to validate using simulation analysis. | Jin et al. [187] |
| 17. | Transmission line | Incipient fault diagnosis in transmission lines using feature extraction algorithm and leakage current signals harmonics using qualitative trend analysis and Naive Bayes classifier. | Silva et al. [188] |
| 18. | Transmission line | Fault diagnosis of dynamic and hybrid domains of transmission line to identify the prominent fault for both persistent and intermittent faults using BN and arithmetic circuits. | Ricks and Mengshoe [168] |
| 19. | Continuous stirred tank heater process | Incipient sensor fault detection and isolation in multimode processes using conditionally independent Bayesian learning based recursive transformed component statistical analysis. | Shang et al. [174] |
| 20. | Water tank system | Fault tree analysis is used to identify the failure component and BN model is used to make relationships between cause and symptom of water tank to diagnose the fault and their root causes. | Lampis and Andrews [43] |
| 21. | Wastewater treatment processes | Fault diagnosis of wastewater treatment plant using expert system, rule-based, and probabilistic approaches of BBN. | Chong and Walley [95] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 22. | Air handling unit | Fault diagnosis of fans, dampers, ducts, filters and sensors of air handling unit using BN model to make causal relationship between fault, fault evidence, and additional information to identify the faults. | Zhao et al. [170] |
| 23. | Air handling unit | Fault diagnosis of air handling units using BN model to analyze the observed behavior and comparing with generated patterns in various faulty conditions. | Najafi et al. [189] |
| 24. | Rotating machine | Identification of root cause of process monitoring variations to diagnose a state of tool wear, workpiece hardness, and a stock size dimensional variation using BN. | Dey and Stori [190] |
| 25. | Hydroelectric generation system | Construction of the BN using expert experiences and vibration data to diagnose the fault and Bayesian inferences are used to develop a CBM policy. | Xu et al. [191] |
| 26. | Hot forming process | Fault diagnosis of hot forming process using $T^2$ statistic and multivariate control charts in a BN. | Verron et al. [30] |
| 27. | Vehicle fault | Multiple symptoms oriented troubleshooting strategy is used to construct a BN model using the infotainment system as a case study to identify root causes of fault occurrence. | Huang et al. [192] |

### 2.6.3.2 *Regression-based fault analysis*

Regression-based fault analysis is a form of predictive modeling technique used to find the relationship between a dependent and an independent variable. It is very sensitive to the outlier. Regression-based methods are used to develop the benchmark for fault detection and handle higher dimensionality of the data set [193], [194]. It requires a large dataset because maximum likelihood estimates are less powerful at small dataset than the ordinary least square method. The regression-based fault analysis approach is categorized into linear regression and logistic regression. The linear regression model is used when the data is continuous, and logistic regression is used when the collected data is discrete in nature. The regression-based fault analysis approach is used in various applications, such as the power transmission line [195], air handling unit [196], aircraft engines [194], [197], and non-stationary processes [198]. The limitation of regression-based fault analysis is that it has a very lengthy and complicated calculation procedure. The functional relationship established between any two or more variables of the system, based on some limited data, does not give an accurate result if more and more data are taken into consideration.

#### *2.6.3.3 Classification based fault analysis*

Classification based fault analysis can be used to classify the output variable into different classes, i.e., normal condition, fault type1, fault type 2, and so on [199], [200]. Three classification algorithms are mostly used in fault classification, such as SVM, HMM and ANN.

*Support Vector Machine*

The SVM based fault analysis is used for pattern recognition for building linear decision boundary to distinguish between two classes, and it can be extended to distinguish among multiple classes ($n$ groups/classes) [44], [201]–[205]. The general idea behind the model is to maximize the shortest distance between the data points and the separating hyperplane; this is often called the functional margin [45], [206], [207]. The SVM based fault analysis is also used to solve both classification and regression problems in industrial applications [207]. The SVM based fault analysis is used in various applications, e.g., internal combustion engine [208], satellite power system [209], solid oxide fuel cell system [210], rotating machine [204], gas-turbine [205], induction motor [44], [129], DC motor [211], chiller fault [202], [206], [212], [213], three-phase electrical system [214], nuclear power plant [45], machine bearing [215], [216], rotating machinery [217], gearbox faults [218], reciprocating compressors [219], and air handling unit [220]. The SVM based fault analysis approach is time-consuming, and it is not suitable for large data sets. The SVM does not perform very well when the target classes are overlapping, and dataset is noisy. Another limitation of application of the SVM based fault analysis is that it underperforms when the number of features for each data point exceeds the number of training data samples.

*Hidden Markov Model*

The HMM is a statistical model-based on the principle of the Markov chain theory. The HMM is widely used for speech recognition systems [46]. Its application has been extended to fault detection, fault diagnosis, and prognostic activities for machining processes as well as estimating the remaining useful life of the system [221]–[223]. The HMM is used in the fault analysis of various applications, such as bearing fault [222]–[226], cutting tool of vertical drilling machine [221], synchronous motors [47], [227], and helicopter main gearbox [228], and bearing of rotating machine [229]. The HMM based fault analysis cannot express the dependencies between the hidden state and it is not practical to represent multiple overlapping features and long-term dependencies. The HMM is unable to capture the high order correlation among the parameters of the system. The HMM often has a large number of unstructured parameters; hence it needs a large number of training data set to give a better result.

### *Artificial Neural Network*

The ANN is a non-statistical quantitative approach used in fault analysis of both process monitoring and historical data based linear and nonlinear system for industrial applications [32]–[35], [230]. The process monitoring based ANN approach is divided into two categories: residual-based fault analysis and pattern recognition based fault analysis [231]. The residual-based fault analysis is used to compare the output of the actual behaviour of the system into model behaviour [232], [233]. The process monitoring and fault diagnosis approach for syntactic pattern recognition is used to identify the trend of noisy process data using the backpropagation-based neural network [36], [56], [234]. Further, ANN based fault analysis approach is divided into two categories: classification based ANN approach and regression-based ANN approach [235]. The ANN based fault analysis used in various industrial applications is presented in Table 2.1.

Table 2.2 ANN based fault analysis approach of various industrial applications

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 1. | Internal combustion engine | Fault diagnosis system of internal combustion engines using wavelet packet transform and ANN to detect and identify the fault. | Wu and Liu [236] |
| 2. | Marine diesel engine | Identification of faults to develop a belief rule-based inference model on evidential reasoning rule for marine diesel engines. | Xu et al. [237] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 3. | Bearing and gear fault of motor | The fault diagnosis of motor bearing and gear using vibration signal using convolutional neural networks and extreme learning machine. | Chen et al. [238] |
| 4. | Gearbox of electric motor | Vibrations features used in fault diagnose of gearbox using ANN and genetic algorithm. | Hajnayeb et al. [57] |
| 5. | Gearbox fault | Vibration experiment used in intelligent fault diagnosis of a gearbox using deep convolutional neural networks model and SVM. | Qiu et al. [239] |
| 6. | Bevel gearbox | Vibration signals is used for incipient gearbox fault diagnosis using discrete wavelet transform for feature extraction, and ANN is used for classification. | Saravanan and Ramachandran [240] |
| 7. | Worm gears | Condition-based fault diagnosis of worm gears using thermal analysis and MLP neural network. | Waqar and Demetgul [241] |
| 8. | Rolling element of bearing | Intelligent fault diagnosis of rolling element bearings using deep distance metric learning and deep neural network clustering algorithm. | Li et al. [242] |
| 9. | LHD machine | Time to failure data is used to estimate the reliability of the load-haul-dump machine using a genetic algorithm-based ANN. | Chatterjee and Bandopadhyay [243] |
| 10. | Induction motors | Fault detection and isolation of induction motor using advanced neural network, genetic algorithms, and expert system. | Betta et al. [244] |
| 11. | Rotating machinery | ANN used for fault prediction of rotor-bearing system of rotating machinery using vibration signals. | Vyas and Satishkumar [58] |
| 12. | Rotating machine | Sound emission signal used for gear fault identification of rotating machinery using wavelet transform and ANN. | Wu and Chan [245] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| 13. | Synchronous generator | Magnetic flux is used for internal faults localization of salient-pole synchronous generator using ANN. | Yaghobi et al. [231] |
| 14. | Wind turbine | Condition monitoring studies used for fault diagnosis of gearbox and bearing fault of wind turbine using ANN. | Biswal and Sabareesh [246] |
| 15. | Steam turbine | ANN is used for fault detection and BN is used for fault isolation and identified the root cause to take a decision for maintenance of steam turbines. | Karlsson et al. [247] |
| 16. | Gas compressor | Fault classification of the univariate and multivariate dynamic systems of gas compressor using weightless ANN. | Oliveira et al. [232] |
| 17. | Sugar Factory | Fault detection, isolation, and fault identification of sugar evaporation process using real-time recorded data using ANN. | Patan and Parisini [248] |
| 18. | Pneumatic system | Fault identification of pneumatic systems collecting data from eight different sensors using ANN. | Demetgul et al. [55] |
| 19. | Induction motor | Fault detection and diagnosis of bearing faults in induction motors based on vibration signals using convolutional neural networks. | Suh et al. [98] |
| 20. | HVAC system | Multiple sensor fault diagnosis of heating, ventilation and air-conditioning (HVAC) system using Auto-Associative neural network. | Elnour et al. [249] |
| 21. | Double lap bolted joints | Prediction of failure stages of double lap bolted joints using finite element analysis and ANN. | Atta et al. [250] |
| 22. | Industrial equipment | Forecasting faults of anode production using ANN classifiers to detect the pattern change 5– | Kolokas et al. [251] |

| Sl. No. | Application | Description | Authors |
|---|---|---|---|
| | for anode production | 10 min before a occurrence of major breakdown of the system. | |

## 2.7 Summary

The comprehensive literature survey of fault analysis methods used for solving various industrial problems are discussed in this chapter. In this chapter, the research works conducted on dragline to optimize the downtime and prevent the occurrence of failure are also described. The research work proposed in this thesis for fault analysis uses FMECA for identifying the risky components of dragline, followed by artificial intelligence based fault analysis. In the literature, three types of fault analysis approaches have been explained. After identifying the critical components of dragline, suitable fault analysis approach should be used to reduce their failure and consequences. The supervised learning of BN model is highlighted for inference based fault analysis of dragline, and ANN model is discussed briefly for classification based fault analysis.

In inference based fault analysis, BN is described as the most powerful tool for fault analysis to explain the causal relationship between multi-layer problems, which uses probability statistics in complex fields in reasons the uncertainty. The BN based fault analysis also provides a good prediction accuracy, avoiding overfitting of data even with rather small and incomplete sample sizes [42]. The BN model is also used to create 'expert system' including expert elicitation and easily combine with decision analytic tools, discrete and continuous monitoring data in the complication system [13], [42]. The historical cause, symptom, and fault data of drag system of dragline is used to analyse the complex causal relationships between them and to derive fault inference for real-time applications in maintenance.

Unlike many fault prediction techniques, ANN based fault analysis does not impose any restrictions on the input variables (e.g., distribution, sample size, linear and complex relationship) and can better model heteroskedasticity i.e., data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data. Hence the ANN model is applicable when the data volatility is very high, and it can work with incomplete knowledge of the system. Therefore, in this thesis, two ANN models, such as cause to symptom and symptom to faults models have been used to find out the nonlinear relationships between the causal variables of drag system.

Though there are many data analysis methods explained in this chapter, the literature survey on fault analysis is mainly focused on inference based BN and classification based ANN with their industrial applications.