Chapter 6

Bayesian Network Based Fault Analysis

6.1 Introduction

The fault analysis of drag system using BN model has been presented in this chapter. Initially, the basic topology of BN model is discussed. After that the methodology used in fault analysis of the drag system is also explained that consists of conflict analysis, fault inference, fault identification and sensitivity analysis. The construction of BN model, estimation of CPT, and identification of threshold limit value of degree of detection of fault is also presented in this chapter. The validation of BN model using conflict analysis and sensitivity analysis is also presented. Finally, the results of the BN model and discussion on fault identification and fault inference through sensitivity analysis have been discussed.

6.2 Topology of BN Model

The BN probabilistic graphical models, also known as belief networks, represent joint probability distributions of sets of random variables with conditional dependencies used as a predicting tool based on Bayesian statistical theory of the supervised machine learning algorithm. A Bayesian network is defined by two components such as structure and parameters. The structure is presented in the form of a directed acyclic graph that consists of a set of nodes and arcs [179]. Generally, nodes represent variables. Each variable has several possible states (e.g., Yes or No; Low, Medium or High; 0 or 1). The arc is a connecting link between the variables and the direction of arc presents the probabilistic dependences between the variables. The parameter of the BN model presents the prior probability of each root node for each state and the CPT of each child node given parental states. For the construction of BN, first generate the influence diagram to describe the system structure and parameters from the collected historical data. The relationship between cause-symptom-fault can be constructed using the CPT of BN, which can be used to predict the fault. The CPT is used to automatically estimate the probability from the collected data and to make the causal relationships between parent node and child node [198], [262], and it has an advantage that it can be regularly updated to generate sufficient information about the health/condition of the system when the new evidence is observed. For discrete cases, CPT shows the probability of various states of a node in accordance with the configuration of a single variable or multiple variables with respect to the parent-child relationship [7], [263]. For fault analysis of drag system, three level structure of BN is formulated: root level, intermediate level, and leaf level. The root level is composed by control variables, the intermediate level is constituted by process characteristics and leaf level is formed by product parameters.

In the construction of the BN model, when number of variables increases, the Bayesian reasoning process increases exponentially. In order to solve the complexity of the calculation of joint probability distribution, there are three independent assumptions [171]. The first assumption is that all the root nodes in the BN are independent of each other. In this study, three sets of variables such as $X = (C_1, C_2, \ldots, C_p; S_1, S_2, \ldots, S_q; F_1, F_2, \ldots, F_r)$ denoted as cause, symptom, and fault, respectively are considered; where p is the number of cause nodes denoted as $C_1, C_2, \ldots, C_p$; q is a number of symptom nodes denoted as $S_1, S_2, \ldots, S_q$; r is the number of fault nodes denoted as $F_1, F_2, \ldots, F_r$ and n is the total number of nodes ($n = p + q + r$) in the BN model. For example, in Figure 6.1, the BN model having three level parameters is presented, such as the root level, the intermediate level and leaf level.

There are two root nodes (cause nodes), and the probability relationship between the root nodes is defined as $P(C_1C_2) = P(C_1)\,P(C_2)$. Secondly, the relationship between two nodes having one or more than one common immediate parent nodes, and there is no direct arrow between them; they are defined as conditionally independent of each other given the states of their immediate parent nodes. For example, C_1 is the parent node of S_1 and S_2, and thus they are conditionally independent of each other given C_1, i.e., $P(S_2 \mid C_1 S_1) = P(S_2 \mid C_1)$ (Figure 6.1). The third assumption refers to non-root node, when the states of its immediate parent nodes are given, then the relationship is defined as a non-root node which is conditionally independent of its non-immediate parent nodes. For example, when the intermediate nodes having S_1 and S_2 are given, then F_1 is independent of C_1 and C_2, i.e., $P(F_1 \mid C_1C_2S_1S_2) = P(F_1 \mid S_1S_2)$ (refer Figure 6.1). These assumptions made the calculations easy, simplified the inference analysis, and significantly reduced the calculation of prior probabilities of each node.

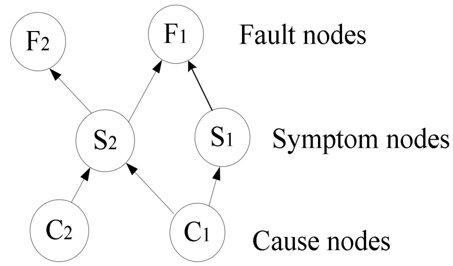


Figure 6.1 Schematic representation of a three-layer BN model

The general equation for the calculation joint probability distribution can be given as a product of the specified conditional probability as presented in Eq. (6.1) [7], [173]:

$$P(X) = P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents\ (\mathrm{X}_i)) \tag{6.1}$$

where $X = X_1, X_2 \ldots, X_n$ is a set of variables in the BN model and n is the number of variables.

The joint probability distribution for a given BN model can be calculated as given in Eq. (6.2) (see Figure 6.1).

$$P(C_1, C_2, S_1, S_2, F_1, F_2) = P(C_1)P(C_2)P(S_1 \mid C_1)P(S_2 \mid C_1C_2)P(F_1 \mid S_1S_2)P(F_2 \mid S_2)) \tag{6.2}$$

where $C_1, C_2, S_1, S_2, F_1, F_2$ are set of variables in the given BN model (see Figure 6.1) representing the cause nodes (C_1, C_2) symptom nodes (S_1, S_2) and fault nodes (F_1, F_2) respectively and the total number of nodes are six. With the help of joint probability distribution, the occurrence of the fault, i.e., $F_2 = 1$ can be calculated using Eq. (6.3) (Figure 6.1).

$$P(F_2=1) \;=\; \sum_{C_1C_2S_1S_2F_1} P(C_1, C_2, S_2, F_2=1, S_1, F_1) \tag{6.3}$$

Further calculating the joint probability of occurrence of fault $F_2 = 1$, when the evidence (symptom) exceeds the threshold limit, i.e., $S_1 = 1$ is given in Eq. (6.4) [7]:

$$\begin{aligned} P(F_2=1 \mid S_1=1) &= \frac{P(F_2=1, S_1=1)}{S_1=1} \\ &= \frac{\sum_{C_1C_2S_2F_1} P(C_1, C_2, S_2, F_2=1, S_1=1,\ F_1)}{\sum_{C_1C_2S_2F_1F_2} P(C_1, C_2, S_1, S_2, F_2=1, F_1)} \end{aligned} \tag{6.4}$$

The limitation of the proposed BN model is that it performs well only when the network structure is already given and it cannot analyse the fault when the relationships are cyclic. In the case study of the drag system of dragline, 16-node BN is considered having cause, symptom, and fault nodes, X= (C_1, C_2, ... , C_6; S_1, S_2, … , S_6; F_1, F_2, … , F_4). In the BN, there are six cause nodes identified for drag system C = (BLB, IC, LC, OL, OH, DC), six symptom nodes S = (CF, VF, VUS, LL, TF, Sp), and four fault nodes F = (BBF, DCP, SBP, ID).

6.3 BN Methodology

For the development of 3-layer BN model, collected training dataset of cause, symptom and fault (Table 5.1) are fed into the Netica software and the causal relationships between the nodes are established. The Netica uses the fastest known algorithm for exact general probabilistic inference in a compiled belief network. The prior probability of each node in the BN model is calculated. During dragline operation, when the evidence of one or more symptoms is observed to exceed the threshold limit, it can be fed to the BN model formulated based on historical data and the CPT will be updated. The evidence of symptoms can be identified either through the sensor feedback or visual inspection by the operator. Based on the prior probability and likelihood of observed evidence, the posterior probability of fault is defined. When the set of evidence (e.g., symptoms) are observed, the BN model can be updated to identify the fault types (e.g., catastrophic, degraded, or intermittent fault). The maintenance support system used decides the suitable maintenance interval based on the fault type which can prevent the occurrence of failure and optimize the downtime. Moreover, the sensitivity analysis is used to identify the root causes and symptoms that most influence the targeted fault, which must be quantified based on the given evidence during maintenance.

The description of the BN model and corresponding condition-based maintenance action plan is shown in Figure 6.2. The maintenance action plan utilizes the results of the condition monitoring to provide users with useful information such as the possible maintenance schedules depending on the accumulated previous data. After maintenance of the dragline, the maintenance action undertaken to reestablish a running system, and the corresponding cause, symptom and fault data are stored. The stored data can be pre-processed and fed in the BN model. When the equipment information is changed through equipment maintenance actions, the configuration management function updates and stores the information through CPT [7], [13].

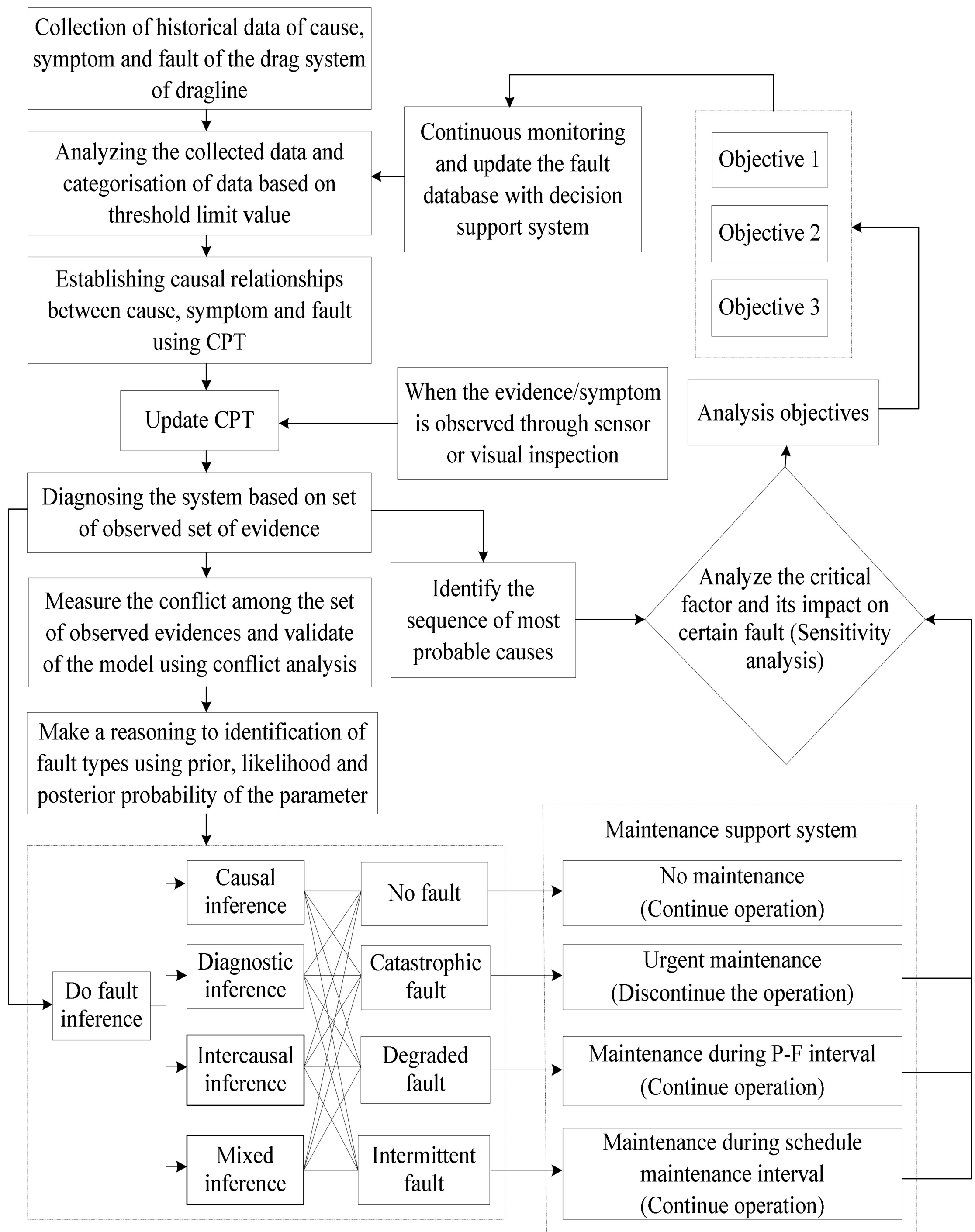


Figure 6.2 Decision flow chart for fault analysis of the drag system

6.3.1 Model validation through conflict analysis

Model validation gives the significant aspect to the real-time fault diagnosis methodology because it provides reasonable confidence to diagnostic results [182], [264]. Notably, a conflict may occur when two or more evidence is provided as input into the BN by sensors or visual observation by an operator. The evidence-driven conflict analysis is necessary to detect possible conflicts among a set of evidences in the BN model. That is, conflict analysis can be used for conflict detection as well as for model validation, which improves reasonable confidence of diagnostic results. A conflict measure is used to demonstrate possible conflict, when the product of the probabilities of the individual pieces of evidence $P(X_i)$ is larger than the joint probability of the evidence $P(X)$ [182], [265], [266]. For the given set of evidence, under the assumption that pieces of evidence are positively correlated, the relationship is presented in Eq. (6.5).

$$P(X) > \prod_{i=1}^{n} P(X_i) \tag{6.5}$$

The general conflict measure is given in Eq. (6.6) [182], [267], [268].

$$Conflict\ (x) = Conflict\ (X_1, X_2, \ldots, X_n) = \log \frac{\prod_{i=1}^{n} P(X_i)}{P(X)} \tag{6.6}$$

The negative value of the conflict (x) indicates that no conflicting evidence exists, and the developed model is correct [182]. The positive value of the conflict measure demonstrates a possible conflict of evidence. Higher the conflict measure, the greater is the discrepancy between the BN model and the evidence. This discrepancy may be due to errors in the data or it just may be a rare case. If the conflict is due to flawed data, it is possible to trace the conflicts [267]. Thus, this evidence-driven conflict analysis can be performed for validation of the complete BN based fault diagnosis models [182]. This measure implies that positive values of conflict measure (x) indicate the positive correlation between sets of evidences and therefore, represent an incorrect model. When the conflict measure is zero, it means it is independent of the observed evidence.

6.3.2 Fault inference

Fault inference is a process of drawing conclusions based on the evidence (cause/symptom/ fault) and its direction of reasoning. The fault inference of BN model updates CPT based on the prior knowledge (prior probability) with subsequent available evidence based on the observed data [13]. The probabilistic reasoning on BNs works in the framework of Bayesian statistics and focuses on the computation of posterior probabilities of any event, called probabilistic inference [40], [262]. There are four types of fault inferences used in BN, as shown in Figure 6.3. In diagnostic inference, also called backward inference, the inference direction is from their child to parent node that estimates the posterior probability of the parent node (query) from the observed child node (evidence) [13]. Causal inference is also called forward inference since the inference direction proceeds from the parent to their child, i.e., it estimates the posterior probability of a child node (query node) of the observed parent node (evidence) [13]. The intercausal inference is used for investigating the complex relations either between causes and a symptom, or between symptoms and a fault [267]. The mixed inference is the combination of rule-based reasoning and case-based reasoning, and it is applied to the intellectualized design. In the case study of the drag system of dragline, four BN inference types (diagnostic, causal, intercausal, and mixed inference) are studied, and their results are interpreted for analyzing faults.

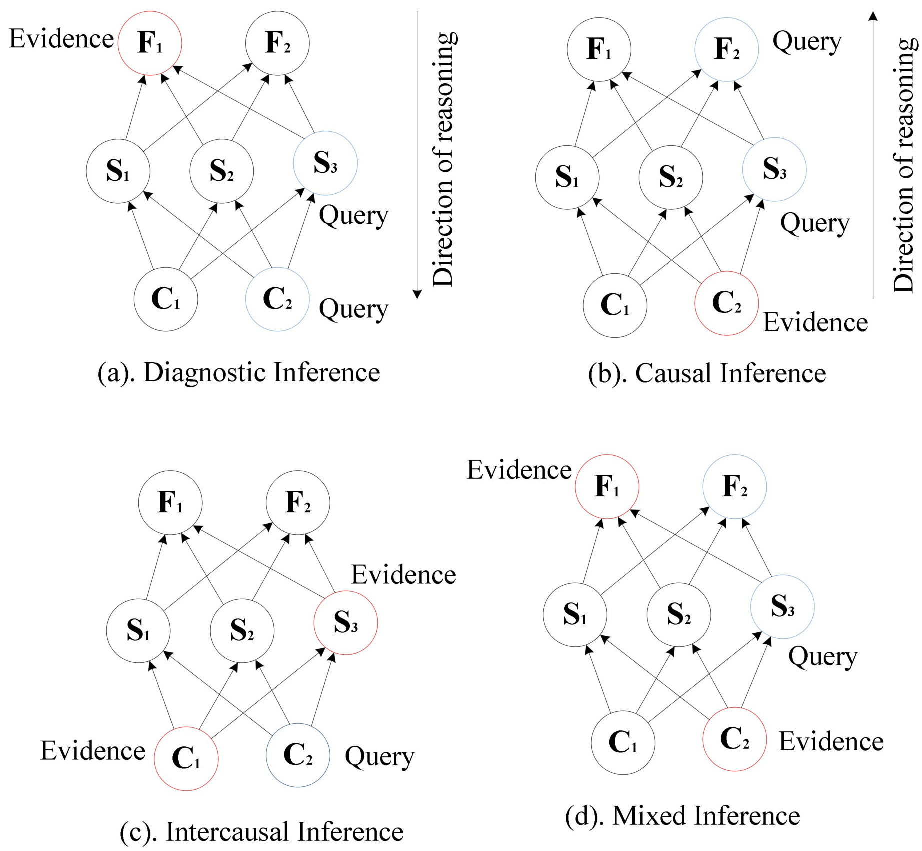
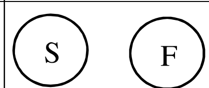
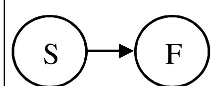
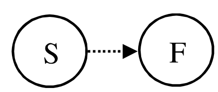
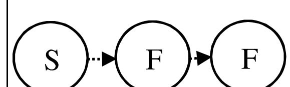


Figure 6.3. Types of fault inference (Source: [13], [267])

6.3.3 Fault identification

After the fault has been detected and isolated, the fault identification step is left to the user's technical experience to determine the size, type, and nature of the fault [13], [269]. The use of threshold values in the identification of fault types, e.g. catastrophic, degraded, or intermittent faults are presented in Table 6.1 [13]. The catastrophic faults are destructive in nature that leads to a large fault in small duration and cause transients behaviour of the system [12]. The degraded faults occur slowly due to gradual deterioration of components [82]. The intermittent faults occur at intervals and these are mostly irregular in occurrence [85]. Based on the occurrence probability of fault, make reasoning to identify the fault types, such as no fault, catastrophic fault, degraded fault or intermittent fault for given evidence as presented in Table 6.1.

Table 6.1 Condition-based fault type identification and its symbolic presentation [13], [170], [270]

Fault type	Condition	Symbolic presentation	Definition
No fault	If • $P(F_i = 1) > P(F_i = 1 \mid S_i = 1)$		No fault
Catastrophic fault (Abrupt Fault)	If • $P(F_i = 1 \mid S_i = 1) > \alpha$ • $P(F_i = 1 \mid S_i = 0) < \beta$ • $P(F_i = 1 \mid S_i = 1) - P(F_i = 1 \mid S_i = 0) > \varepsilon$ • $P(F_i = 1 \mid S_i = 0) < P(F_i = 1) < P(F_i = 1 \mid S_i = 1)$		These are large faults occurring in small duration and they cause transients in system behaviour.
Degraded fault (Incipient Fault)	If • $P(F_i = 1 \mid S_i = 1) > \alpha$ • $P(F_i = 1 \mid S_i = 0) > \beta$ • $P(F_i = 1 \mid S_i = 1) - P(F_i = 1 \mid Si = 0) < \varepsilon$ • $P(F_i = 1 \mid S_i = 0) < P(F_i = 1) < P(F_i = 1 \mid S_i = 1)$		These faults occur slowly due to gradual deterioration of component.
Intermittent or persistent fault (Periodic Fault)	If • $P(F_i = 1 \mid S_i = 1) < \alpha$ • $P(F_i = 1 \mid S_i= 0) > \beta$ • $P(F_i = 1 \mid S_i = 1) - P(F_i = 1 \mid S_i = 0) < \varepsilon$ • $P(F_i = 1 \mid S_i = 0) < P(F_i = 1) < P(F_i = 1 \mid S_i = 1)$		These faults occur at intervals and these are mostly irregular in occurrence.

In Table 6.1, α, β, and ε is the degree of fault detection called threshold values decided by considering some specific fault occurrence cases in combination with the experts' opinion and available maintenance records [13], [40], [270]. The term $P(F_i = 1|S_i = 1)$ is representing the probability of occurrence of fault after evidence is observed while the term $P(F_i = 1|S_i = 0)$ representing the probability of occurrence of fault before evidence is observed. The term $P(F_i = 1)$ represents the likelihood of the occurrence of fault.

6.3.4 Sensitivity analysis of BN model

In BNs, the sensitivity analysis can be used for verifying the correctness of parameters, and to understand whether more precision in estimating them would be useful [271], [272]. During maintenance, when the limited maintenance resources are available, it is needed to know the root cause that mostly influences the occurrence of fault and failure of the dragline. To fulfill the object, a sensitivity analysis in BN model can be conducted. Two types of sensitivity analysis are used, in one-way sensitivity analysis, i.e. the case in which a single parameter is a varied and corresponding change in posterior probability of occurrence of target fault must be quantified [13]. On the other hand, in multiple-way sensitivity analysis, two or more than two parameters varied are simultaneously.

The sensitivity analysis investigates the effect of small changes in the numerical value of input parameter (i.e., probabilities) and identifies the accordance change in the numerical value of the output parameter (e.g., posterior probabilities). The results show that the changes are the probability of occurrence of a fault and identify the more sensitive parameter according to available evidence of symptoms. Highly sensitive parameters of the BN model affect the performance of the dragline significantly; thus, requiring more effort to analyze and understand these parameters and to obtain more accurate values with a good understanding of the BN. On the other hand, less sensitive parameters of the BN model are very difficult to identify. If the sensitivity value is less than one, then a small change in the likelihood value has a minimal effect on the result of the posterior output of the hypothesis. The results of sensitivity analysis would also provide information about the "robustness" of the model parameters that can help the maintenance engineer during the decision-making process.

In the sensitivity analysis, three types of parameters are used to investigate the relationship between prior and posterior probability such as interesting parameter, evidence, and parameter understudy, as shown in Figure 6.4 [13], [273]. The first type of variable is the target variable (interesting node) in diagnostics model of the BN, which represent the uncertain events which are unobservable but of interest to identify (i.e. fault node) to improve the performance of HEMM. The second type of observable variables provides the useful information, with the help of sensors or experts' opinion that can be put in the BN model and the CPT can be updated, which is known as evidence (i.e. symptom or fault nodes). The third type of variable is the investigating variable or parameter under study to identify (e.g. cause or symptom) that are responsible for the occurrence of a fault.

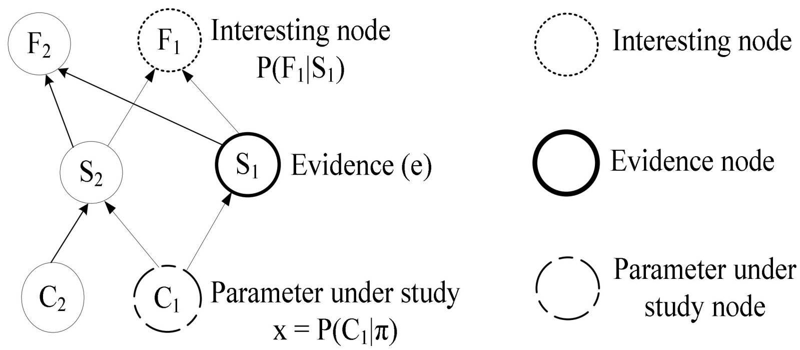


Figure 6.4. Sensitivity analysis in BN

The relative influence of a cause or a symptom on the target fault must be quantified, perhaps by focusing on the more important factors within limited maintenance resources. To identify the influence of the different points of view, three objectives are proposed. To achieve objectives of sensitivity analysis, two values are measured; the absolute sensitivity value (M_1) and the difference between maximum and minimum of the sensitivity function value (M_2). In Figure 6.4, the probability of interest given by the evidence (e) i.e., P($F_1|S_1$), and the functional relationship between the probability of interest and the conditional probability x = p($C_1|\pi$) can now be estimated, and π is a combination of parent values of C_1. The sensitivity function value can be written, as mentioned in Eq. (6.7) [274].

$$f_{P(F_1|S_1)}(x) = \frac{ax+b}{cx+d} \tag{6.7}$$

In Eq. (6.7), a, b, c and d are constants. The parameter sensitivity function of x given for evidence e is calculated as the partial derivative with respect to x obtained by differentiating Eq. (6.7). The absolute sensitivity value M_1 is represented in Eq. (6.8) [13].

$$M_1 = \left| f'_{P(F_1|S_1)}(x) \right| = \left| \frac{ad - bc}{(cx+d)^2} \right| \tag{6.8}$$

The maximum M_1 value means that the parameter under investigation has a significant effect on the interesting parameter for the given evidences, and it decides the necessity of careful monitoring of the parameter. Calculating the difference between maximum sensitivity value and minimum sensitivity value, generally denoted as M_2, helps in analyzing the impact of the parameter in accordance with the occurrence of the interesting probability as presented in Eq. (6.9) [13].

$$M_2 = \text{Max}\, f(x) - \text{Min}\, f(x) \tag{6.9}$$

The M_2 value is maximum means that the probability of occurrence of the interesting node with change in parameter probability for given evidence is most sensitive that can be helpful to analyze the fault. The most sensitive parameter having maximum M_2 value for given evidence and interesting parameter is calculated in the training dataset and the same parameter for a given evidence and interesting parameter used to calculate the M_2 value for validation dataset. The accuracy of the BN model is defined in the term of percentage of error between model constructed from the training dataset and validation dataset as presented in Eq.(6.10). [275]

$$\text{Error (\%)} = \frac{(\text{M}_2\ \text{maximum})\ \text{test dataset} - (\text{M}_2\ \text{observed})\ \text{validation dataset}}{(\text{M}_2\ \text{maximum})\ \text{test dataset}} \times 100 \tag{6.10}$$

Therefore for conducting sensitivity analysis, it is indeed required to identify the combination of three elements that maximize the M_1 or M_2 value for a specific objective. Three objectives of the sensitivity analysis for studying the occurrence of fault (interesting parameter) in the drag system have been explained to understand the role of given evidence and parameter under study.

- Objective 1: In order to reduce the occurrence probability of a fault, when the parameter under study is highlighted; control the parameter probability for a given evidence. This problem can be resolved by finding the parameter probability (x) for the given parameter and the evidence that maximizes M_1.

- Objective 2: In order to find the parameter and evidence that mostly affect the given interesting fault, the evidence as well as the parameter that maximizes the M_2 value is observed.

- Objective 3: When new evidence is available, the impact on the interesting parameter given the previous evidence and parameter under study, there is a need to identify the parameter probability that maximizes the M_2 value.

The development of a real system, a three-axiom-based sensitivity analysis proposed by Jones et al. [257] was used to partially validate the sensitivity of the BN model for individual cases to demonstrate that the proposed methodology is reasonable and it can diagnose the faults appropriately. The three axioms based sensitivity analysis proposed by Jones et al. [257] is presented below:

- Axiom 1. A slight increase/decrease in the prior subjective probabilities of each parent node should certainly result in the effect of a relative increase/decrease of the posterior probabilities of child nodes.
- Axiom 2. Given the variation of subjective probability distributions of each parent node, its influence magnitude to the child node values should be consistent.
- Axiom 3. The total influence magnitudes of the combination of the probability variations from x attributes (evidence) on the values should always be greater than the one from the set of $x - y$ ($y \in x$) attributes (sub-evidence).

6.4 Estimation of CPT

The proposed BN is constructed based on the gathered data of the drag system of the dragline. The faults of the drag system have been identified through threshold limit fault inference based on evidence and their direction of reasoning, and it is then used in sensitivity analysis based on the available information to find the most effective parameter that influences the fault. The estimation of CPT in the BN has been done using the Netica software (Figure 6.5) [256]. The black bars in Figure 6.5 represent the probability of the corresponding state. An arrow indicates the causal relationship either between cause and symptom or between symptom and fault. Results of fault identification and sensitivity analysis can be integrated together, and the CPT can be regularly updated to find out the relevant, effective parameter of each fault type to make suitable CBM strategy [41].

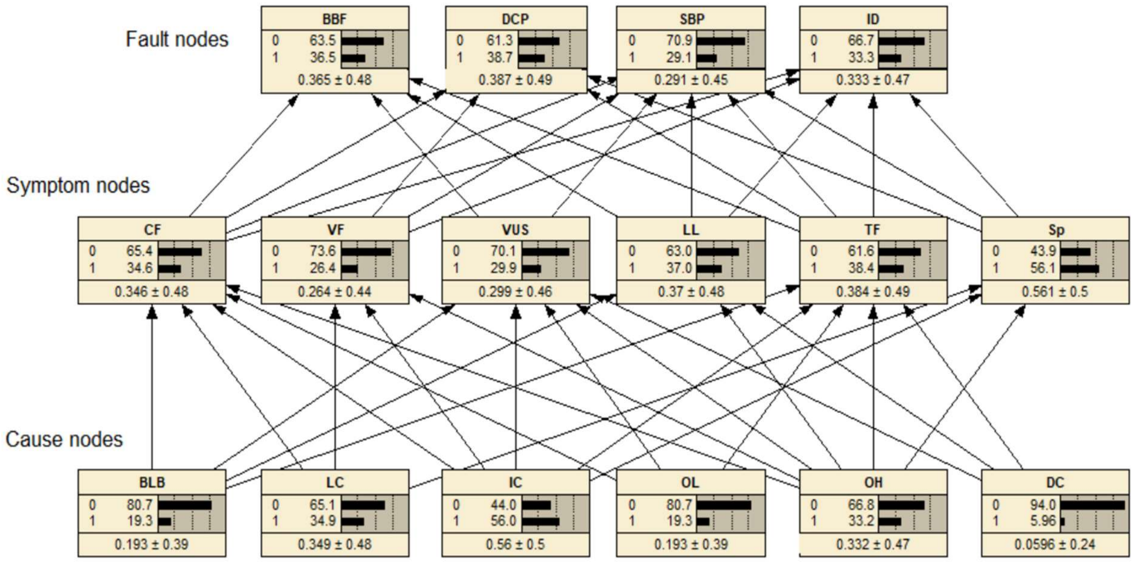


Figure 6.5. Updated CPT of BN for fault analysis of drag system using test dataset

The models satisfy three axioms based sensitivity analysis proposed by Jones et al. [257]. For example, when the first symptom CF = 0 was put in the BN model, the occurrence probability of fault BBF increased from 0.365 to 0.380 and the probability of the fault of the 'fault variable' of remaining faults decreased. After that, the second symptom VF = 0 was put in the BN model and the occurrence probability of fault BBF increased to 0.389. After that, the third symptom Sp = 0 was put in the BN model and the occurrence probability of fault BBF increased to 0.399. After that, the fourth symptom LL = 1 was put in the BN model, the occurrence probability of fault BBF increased to 0.483. After that, the fifth symptom VUS = 1 was put in the BN model and the occurrence probability of fault BBF increased to 0.774. Finally, the last symptom TF = 1 was put in the BN model and the occurrence probability of fault BBF further increased to 0.800. The exercise of increasing each influencing node satisfied these axioms stated in [257], thereby partially validating the model to demonstrate that the proposed methodology and available data used in the model are reasonable. The BN model is further validated using conflict analysis that improved the reasonable confidence of fault diagnosis results; thus proving its efficient sampling and training of the hypothesized BN model.

6.5 Identification of Threshold Limit Value

Before fault identification, several threshold values were decided on the basis of experts' opinion, available data, and characteristics of each fault. The testing dataset was used for fault identification in the drag system, as shown in Figure 6.5. The fault types have been categorized based on the value of the threshold limit of α, β, and ε which denotes the degree of fault detection to categorize the fault types (Table 6.1). The value of a and β is fixed based on the posterior probability of fault after identifying the evidence P(Fi = 1 |Si = 1) and prior probability (before evidence is observed) P(Fi = 1 |Si = 0) respectively. The value of ε is fixed based on the difference between posterior and prior probability of occurrence fault. The value of a = 0.55, β = 0.25 and ε = 0.30 are assumed with the help of experts' opinion, and it was verified using some specific fault occurrence cases available in the maintenance records based on the test dataset. The verification of the threshold limit value of α, β, and ε based on the occurrence of fault based on the available maintenance worksheet data and prior, likelihood, and posterior probability of parameters as presented in Table 6.2.

Table 6.2 Verification of the threshold limit values of α, β, and ε based on occurrence of few faults

EHMR	Previous maintenance record and sensor data of fault occurrence	Fault	Results derived from BN using test dataset			
			$P(Fi = 1 \mid Si = 0)$	$P(Fi = 1)$	$P(Fi = 1 \mid Si = 1)$	$P(Fi = 1 \mid Si = 1) - P(Fi = 1 \mid Si = 0)$
46:00	Sudden increase in temperature above the threshold limit value (TF = 1) and it occurred due to overheating of the drive control system, but the dragline continued to work and its maintenance was done during scheduled maintenance interval. The fault DCP was intermittent in nature.	DCP	0.32	0.387	0.483	0.163
		BBF	0.345	0.365	0.109	−0.236
		SBP	0.312	0.291	0.298	−0.014
		ID	0.341	0.333	0.329	−0.012
326:45	The vibration and unwanted sound were observed (i.e., VUS = 1) and operation of the dragline was continued. Maintenance was done during the scheduled inspection interval due to degraded nature of the occurrence of fault BBF.	BBF	0.234	0.365	0.616	0.382
		DCP	0.378	0.387	0.08	−0.298
		SBP	0.248	0.291	0.232	−0.016
		ID	0.327	0.333	0.02	−0.307
938:15	Increase temperature and sparking were identified (TF = Sp = 1), and urgent maintenance was taken to stop the operation due to catastrophic nature of the occurrence of fault DCP.	DCP	0.238	0.387	0.75	0.512
		BBF	0.353	0.365	0.109	−0.244
		SBP	0.327	0.291	0.188	−0.139
		ID	0.351	0.333	0.313	−0.038
2150:45	Sudden decrease in the lubrication level below the threshold limit value (LL=1), that degraded the performance of bearing and maintenance was done during P-F interval. Dragline operated continuously and the lubricant was filled during schedule maintenance period to optimize downtime.	BBF	0.320	0.365	0.650	0.330
		DCP	0.362	0.387	0.08	−0.282
		SBP	0.264	0.291	0.25	−0.014
		ID	0.318	0.333	0.318	0
3610:30	Sudden sparking was observed (Sp=1), dragline continued to operate and maintenance was done during schedule inspection interval due to intermittent nature of fault ID.	ID	0.299	0.333	0.363	0.064
		BBF	0.376	0.365	0.15	−0.266
		DCP	0.357	0.387	0.08	−0.277
		SBP	0.29	0.291	0.063	−0.227
3711:45	The current feedback and amplitude of the vibration exceeded the threshold limit (CF=VUS=1). The urgent maintenance was carried out to stop the operation due to occurrence of catastrophic fault BBF and degraded fault in drive control system.	BBF	0.257	0.365	0.750	0.493
		DCP	0.320	0.387	0.697	0.377
		SBP	0.191	0.291	0.167	−0.024
		ID	0.244	0.333	0.343	0.099

EHMR	Previous maintenance record and sensor data of fault occurrence	Fault	Results derived from BN using test dataset			
			$P(Fi = 1 \mid Si = 0)$	$P(Fi = 1)$	$P(Fi = 1 \mid Si = 1)$	$P(Fi = 1 \mid Si = 1) - P(Fi = 1 \mid Si = 0)$
3832:15	The current feedback exceeded the threshold limit and sparking were identified (CF=Sp=1) and the operation of the dragline was continued. The maintenance was done during scheduled inspection interval due to occurrence of degraded fault in the drive control system.	DCP	0.303	0.387	0.597	0.294
		BBF	0.390	0.365	0.087	−0.303
		SBP	0.257	0.291	0.233	−0.024
		ID	0.282	0.333	0.273	−0.009
3889:00	The temperature feedback exceeded the threshold limit, lubrication level was below the predefined limit (TF=LL=1). The operation of the dragline was continued. The maintenance was done during P-F interval of the occurrence of fault DCP.	DCP	0.314	0.387	0.483	0.169
		BBF	0.309	0.365	0.270	−0.039
		SBP	0.240	0.291	0.027	−0.213
		ID	0.257	0.333	0.227	−0.030
4165:30	The vibration amplitude and temperature exceeded the predefined limit (VUS=TF=1) and urgent maintenance was taken to stop the operation due to occurrence of catastrophic fault BBF and degraded drive control system.	BBF	0.201	0.365	0.812	0.611
		DCP	0.306	0.387	0.483	0.177
		SBP	0.289	0.291	0.279	−0.010
		ID	0.362	0.333	0.329	−0.033
4378:30	The current feedback exceeded the threshold limit (CF=1) and lubrication level was below the predefined limit (LL=1). The operation of the dragline was continued. The maintenance was done during scheduled inspection interval due the degradation of drive control system and occurrence of fault SBP.	DCP	0.263	0.387	0.697	0.434
		SBP	0.236	0.291	0.585	0.349
		BBF	0.311	0.365	0.190	−0.121
		ID	0.245	0.333	0.179	−0.066

Similarly, when the fault was verified, with respect to the corresponding maintenance actions and the nature of the fault, the threshold limit value of α, β, and ε was verified and decided. The value of $\alpha = 0.55$, $\beta = 0.25$ and $\varepsilon = 0.30$ fulfilled the criteria of identified fault types (Table 6.2).

6.6 Results and Discussions of BN Based Fault Analysis

The results and discussion of the proposed BN model are presented in three sub-sections: fault type identification, accuracy of BN model and sensitivity analysis.

6.6.1 Fault type identification

The threshold limit value and reasoning of fault type identification (refer Table 6.1) are used to identify the fault types (e.g., catastrophic fault, degraded fault, or intermittent fault) based on the available evidence. Six cases are explained.

Case 1: One symptom VUS exceeded the threshold limit and remaining symptoms were within the threshold limit. When the evidence (VUS = 1, CF = VF = LL= TF = Sp = 0) is put in the BN model, the occurrence probability of fault 'BBF' increased from 0.365 to 0.616, as shown in Figure 6.6.

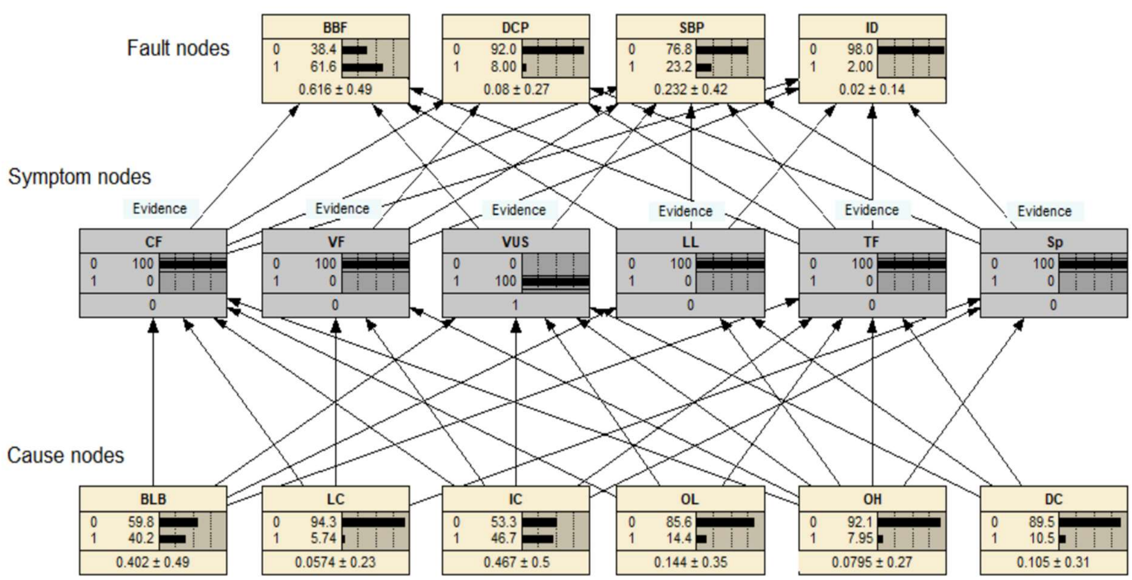


Figure 6.6. Updated BN when symptom VUS is observed in the drag system

The fault BBF is analyzed as per the following observed evidence.

- P(BBF = 1 |VUS = 1, CF = VF = TF = LL = Sp = 0) = 0.616 > a
- P(BBF = 1 |VUS = 0) = 0.324 > β
- P(BBF = 1) = 0.365
- P(BBF = 1 |VUS = 1, CF = VF = TF = LL = Sp = 0)− P(BBF = 1 |VUS = 0) = 0.616− 0.324 = 0.292 < ε

Hence the probability of occurrence of the fault BBF is degraded fault, which means the maintenance engineer can take the maintenance action during P-F interval. The occurrence of posterior probability of remaining faults decreased, i.e., (P(Fi = 1) > P(Fi = 1 |Si = 1). The occurrence probability of fault DCP decreased from 0.367 to 0.080, fault SBP decreased from 0.291 to 0.232 and fault ID decreased from 0.333 to 0.020. Hence they are categorized as 'no fault'. The measured conflict value for the given set of evidence was – 0.161, which revealed free conflict and correctness of the model.

Case 2: When the symptom CF exceeded the threshold limit and the remaining symptoms were within the threshold limit, the occurrence of fault was studied. The evidence (CF = 1, VF = VUS = LL= TF = Sp = 0) was put in the BN model, and it was observed that the occurrence probability of fault 'DCP' increased from 0.387 to 0.697 and fault 'SBP' increased from 0.291 to 0.443, as shown in Figure 6.7.

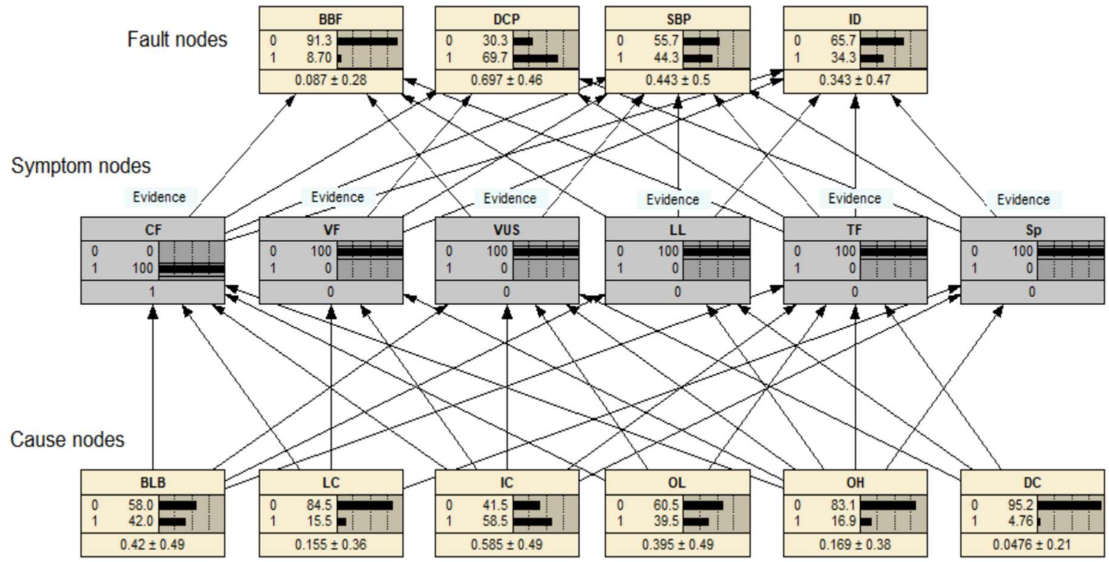


Figure 6.7. Updated BN when symptom CF is observed in drag system

The fault DCP is analyzed as per the following observed evidence.

- P(DCP = 1|CF= 1, VF = VUS = TF = LL = Sp =0) = 0.697 > a
- P(DCP = 1|CF= 0) = 0.319 < β
- P(DCP = 1) = 0.387
- P(DCP =1|CF= 1, VF = VUS = TF = LL = Sp = 0)− P(DCP = 1 |CF= 0) = 0.697− 0.319 = 0.378 > ε

Hence the probability of occurrence of the fault DCP is degraded type, which suggests the maintenance action to be undertaken during P-F interval to minimize the downtime.

The fault SBP is analyzed as per the following evidence.

- P(SBP = 1|CF= 1, VF = VUS = TF = LL = Sp = 0) = 0.443 < a
- P(SBP = 1|CF= 0) = 0.254 > β
- P(SBP = 1) = 0.291
- P(SBP =1|CF= 1, VF = VUS = TF = LL = Sp =0) − P(SBP = 1|CF= 0) = 0.443− 0.254 = 0.189 > ε

Hence the probability of occurrence of the SBP is intermittent fault type. Therefore, the maintenance engineer can take the maintenance action during schedule maintenance interval to minimize the downtime.

The fault ID is analyzed as per the following observed evidence.

- $P(ID = 1 | CF= 1, VF = VUS = TF = LL = Sp = 0) = 0.343 < \alpha$
- $P(ID = 1 | CF= 0) = 0.249 > \beta$
- $P(ID = 1) = 0.333$
- $P(ID = 1 | CF= 1, VF = VUS = TF = LL= Sp = 0) - P(ID = 1 | CF= 0) - P(ID = 1 | CF= 0) = 0.343 - 0.249= 0.094 < \varepsilon$

Hence the probability of occurrence of the fault ID is degraded type, which suggests the maintenance action to be undertaken during P-F interval to minimize the downtime. The occurrence of posterior probability of remaining faults BBF decreased from 0.365 to 0.087. Hence they are categorized as no fault. The measured conflict value for the given set of evidence was – 4.012, which revealed free conflict and correctness of the model.

Case 3: Two symptoms CF and TF exceeded the threshold limit and remaining symptoms were within the threshold limit. When the evidence (CF = TF = 1, VF = VUS = LL = Sp = 0) was put in the BN model, the occurrence probability of fault 'ID' increased from 0.333 to 0.902, as shown in Figure 6.8.

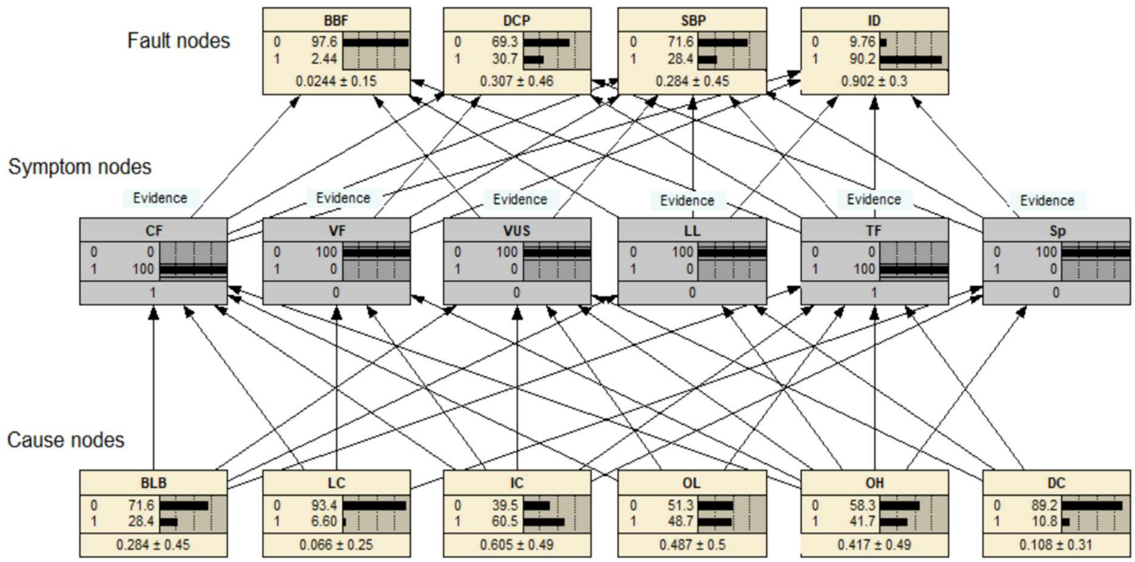


Figure 6.8. Updated BN when symptoms CF and TF are observed in drag system

The fault ID is analyzed as per the following observed evidence:

- P(ID = 1 |CF = TF = 1, VF = VUS = LL = Sp = 0) = 0.902 > a
- P(ID = 1 |CF = TF = 0) = 0.21 < β
- P(ID = 1) = 0.333
- P(ID = 1 |CF = TF = 1, VF = VUS = LL = Sp=0) − P(ID = 1 |CF = TF = 0) = 0.902− 0.210 = 0.692 > ε

Hence the probability of occurrence of the fault ID is categorized as catastrophic fault. When the fault is catastrophic in nature the maintenance engineer should undertake urgent maintenance action and the machine operation in nature should be stopped immediately. The occurrence of posterior probability of remaining faults decreased, i.e., (P(Fi = 1) > P(Fi = 1 |Si = 1). The occurrence probability of fault BBF decrease from 0.365 to 0.024, fault DCP decreased from 0.387 to 0.307 and fault SBP decreased from 0.291 to 0.284. Hence they are categorized as no fault. The measured conflict value for the given set of evidence was – 2.017, which revealed free conflict and correctness of the model.

Case 4: Two symptoms VUS and LL exceeded the threshold limit and remaining symptoms were within the threshold limit. When the evidence (VUS = LL = 1, CF = VF = TF = Sp = 0) was fed as input in the BN model, the occurrence probability of fault 'BBF' increased from 0.365 to 0.650, as shown in Figure 6.9.

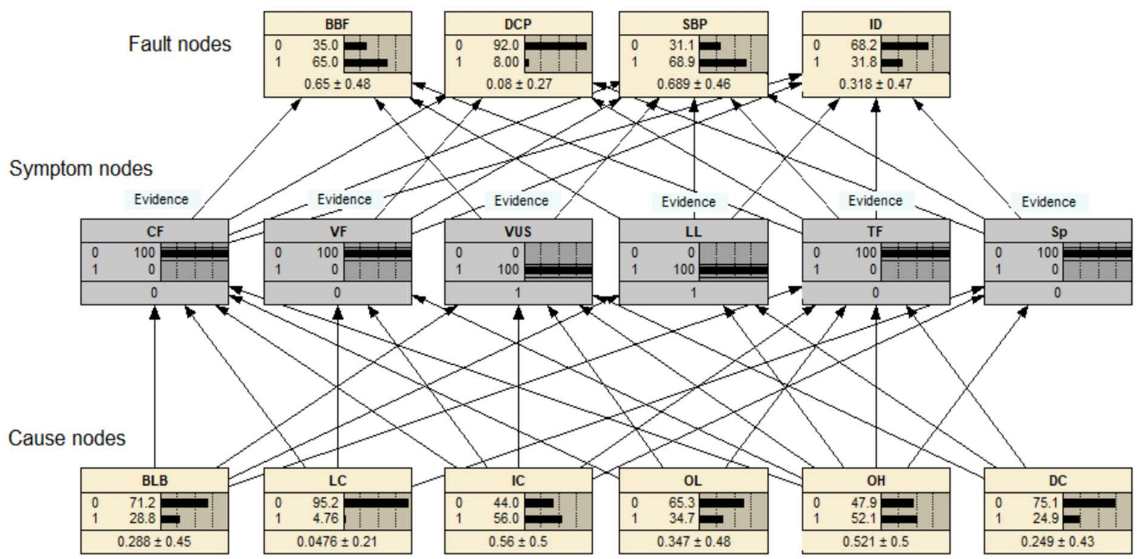


Figure 6.9. Updated BN model when symptoms VUS and LL are observed in the drag system

The fault BBF is analyzed as per the following observed evidence.

- P(BBF = 1 |VUS = LL = 1, CF = VF = TF = Sp = 0) = 0.65 > a
- P(BBF = 1 |VUS = LL= 0) = 0.319 > β
- P(BBF = 1) = 0.365
- P(BBF = 1 |VUS = LL = 1, CF = VF = TF = Sp = 0) − P(BBF = 1 |VUS = LL = 0) = 0.650− 0.319 = 0.331 > ε

Hence the occurrence of the fault BBF is degraded fault. It suggests the maintenance engineer to undertake the maintenance action during P-F interval.

The fault SBP is analyzed as per the following observed evidence.

- P(SBP = 1 |VUS = LL= 1, CF = VF = TF = Sp = 0) = 0.689 > a
- P(SBP = 1 |VUS = LL = 0) = 0.215 < β
- P(SBP = 1) = 0.291
- P(SBP = 1 |VUS = LL =1, CF = VF = TF = Sp =0) − P(SBP = 1 |VUS = LL = 0) = 0.689 − 0.215 = 0.474 > ε

Hence the probability of occurrence of the fault SBP is a catastrophic fault. It suggests the maintenance engineer to undertake urgent maintenance action. The occurrence of posterior probability of remaining faults decreased, i.e., (P(Fi = 1) > P(Fi = 1 |Si = 1). The occurrence probability of fault DCP decreased from 0.367 to 0.080 and fault ID decreased from 0.333 to 0.318. Hence they are categorized as no fault. The measured conflict value for the given set of evidence was – 4.925, which revealed free conflict and correctness of the model.

Case 5: Two symptoms CF and VF exceeded the threshold limit and remaining symptoms were within the threshold limit. When the evidence (CF = VF = 1, VUS = TF = LL = Sp = 0) was put in the BN model, the occurrence probability of fault 'ID' increased from 0.365 to 0.650, as shown in Figure 6.10.

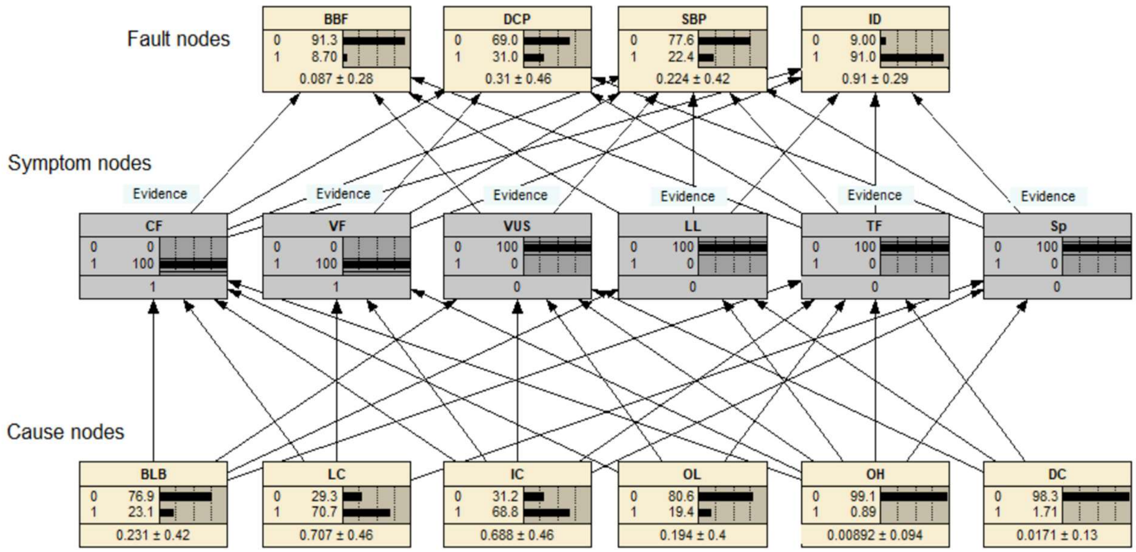


Figure 6.10. Updated BN when symptoms CF and VF are observed in drag system

The fault ID is analyzed as per the following observed evidence.

- P(ID = 1 |CF = VF = 1,VUS = TF = LL = Sp = 0) = 0.910 > a
- P(ID = 1 |CF = VF = 0) = 0.247 < β
- P(ID = 1) = 0.333
- P(ID = 1 |CF = VF = 1, VUS = TF = LL = Sp = 0) − P(ID = 1 |CF = VF = 0) = 0.91 − 0.247 = 0.663 > ε

Hence the probability of occurrence of the fault ID is catastrophic fault, which suggests the maintenance engineer to take urgent maintenance action. The occurrence of posterior probability of remaining faults decreased e.g. (P(Fi = 1) > P(Fi = 1 |Si = 1). The occurrence probability of fault BBF decreased from 0.365 to 0.087, fault DCP decreased from 0.387 to 0.310, and fault SBP decreased from 0.291 to 0.224. Hence they are categorized as no fault. The measured conflict value for the given set of evidence was – 0.628, which revealed free conflict and correctness of the model.

Case 6: Three symptoms CF, TF and Sp exceeded the threshold limit and remaining symptoms were within the threshold limit. When the evidence (CF = TF = Sp = 1, VF = VUS = LL = 0) was incorporated in the BN model, the occurrence probability of fault 'ID' increased from 0.333 to 0.767, as shown in Figure 6.11.

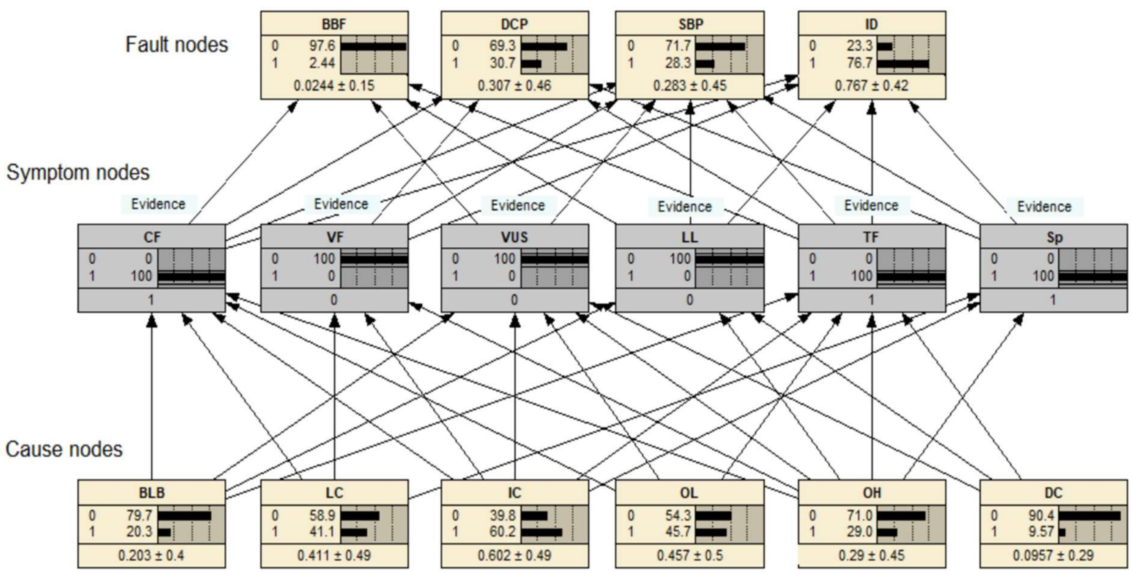


Figure 6.11. Updated BN when symptoms CF, TF, and Sp are observed in drag system

The fault ID is analyzed as per the following observed evidence.

- P(ID = 1 |CF = TF = Sp = 1, VF = VUS = LL = 0) = 0.767 > a
- P(ID = 1 |CF = TF = Sp 0) = 0.103 < β
- P(ID = 1) = 0.333
- P(ID = 1 |CF = TF = Sp 1, VF = VUS = LL = 0) − P(ID = 1 |CF = TF = Sp = 0)

  = 0.767 − 0.103 = 0.664 > ε

Therefore the probable occurrence of the fault ID is a catastrophic fault, and the maintenance engineer should undertake urgent maintenance action in order to prevent the failure of dragline. The occurrence of posterior probability of remaining faults decreased e.g., P(Fi = 1) > P(Fi = 1 |Si = 1). The occurrence probability of fault BBF decreased from 0.365 to 0.024, fault DCP decreased from 0.387 to 0.307 and fault SBP decreased from 0.291 to 0.283. Hence they are categorized as no fault. The measured conflict value for the given set of evidence was – 2.383, which revealed free conflict and correctness of the model. Similarly, the fault types for all possible combinations of cases for given set of evidence can be identified that will help to develop the structure of maintenance strategy.

6.6.2 Identification of critical parameters and accuracy of the model

Three objectives are identified for sensitivity analysis using the test dataset to identify the best combination of the three elements, such as interesting probability, evidence, and parameter under study. The combination which maximizes the M_1 or M_2 value is identified as the most influencing parameter. If the absolute sensitivity value (M_1) is the maximum at the parameter probability value (x) closer to one, the parameter under study has a significant effect on the interesting parameter for given evidences. On the other hand, when maximum M_1 value occurs closer to zero value of x, it can be concluded that it has a lesser effect on the parameter under given evidence. The M_2 value is maximum means that parameter understudy has a significant effect on the interesting parameter. The combination of elements that provides maximum M_2 value in test dataset is identified. The same combination of elements is now studied for finding the observed M_2 value in validation dataset.

6.6.2.1 Sensitivity analysis of fault BBF

In the sensitivity analysis of the fault BBF, it is given that the interesting parameter BBF = 1, evidence VUS = 1, and parameter under study is BLB. In objective 1, when the parameter probability (x), is sought; the maximum value of M_1 is observed as 8.942 at $x = 0$, which means that it has most effect on the parameter under given evidence. The relationship between x and $f(x)$ is shown in Figure 6.12.

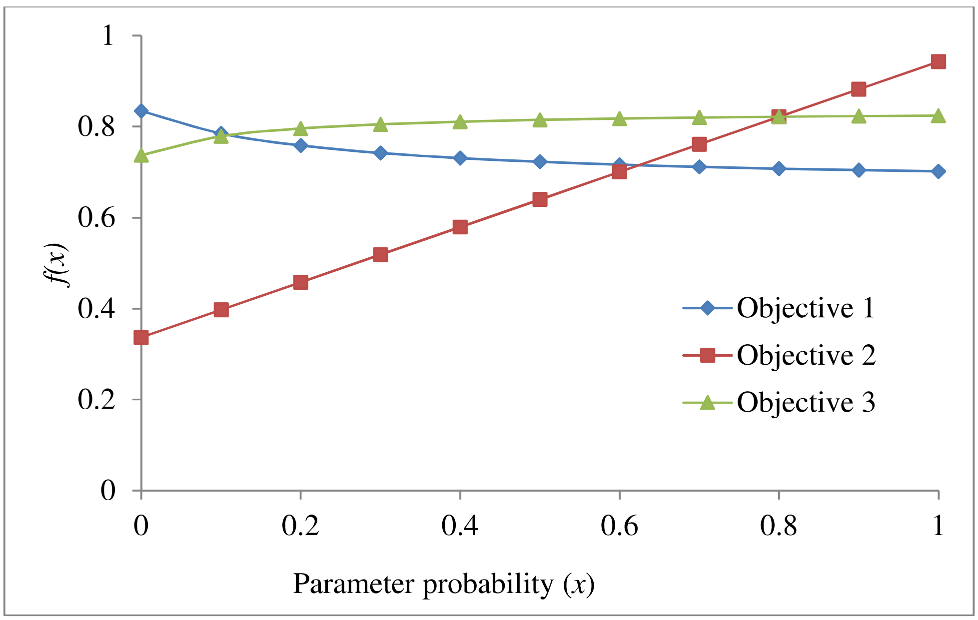


Figure 6.12. Sensitivity function graph of fault BBF for various objectives

In objective 2, when the parameter is sought for given interesting probability BBF = 1, and evidence BLB = 1, that maximizes M_2 value is 0.606 under the parameter, VUS = 1|(BLB = 1, IC = 0, OL = 0, OH = 0, DC = 0), and the relationship between parameter probability x and $f(x)$ is shown in Figure 6.12. In objective 3, for the given the interesting parameter BBF = 1 and evidence VUS =1, and new observed evidence BLB =1 that maximizes M_2 value is 0.087 under the parameter, TF = 1|(BLB = 1, IC = 1, OL

= 0, OH = 1, DC = 0). The relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.12. Hence the parameter becomes more effective when new evidence is observed. The most effective parameter in interesting node BBF has been validated using validation dataset, as illustrated in Figure 6.13. The maximum M_2 value was 0.095 and error was 9.20 %.

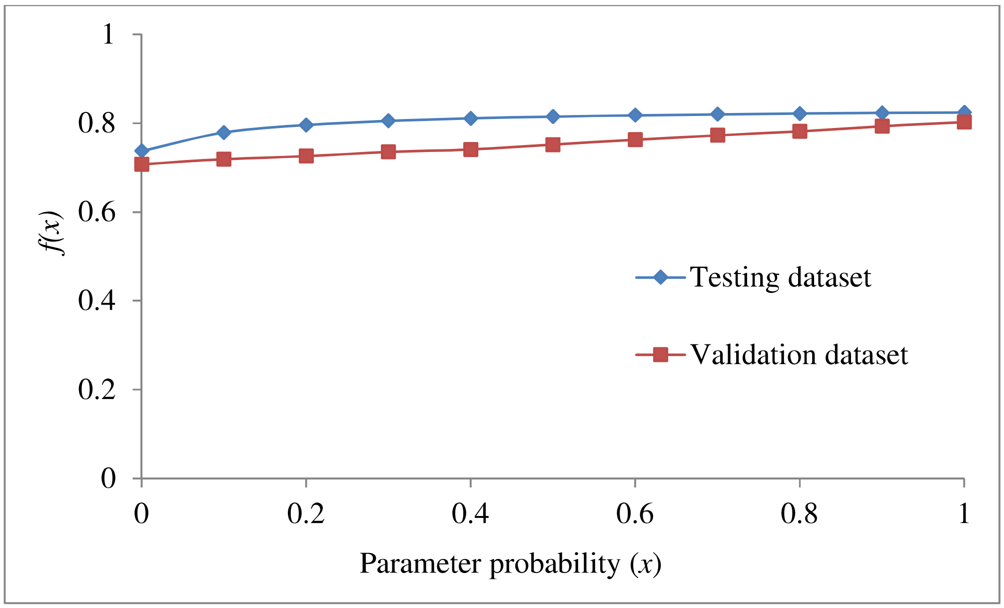


Figure 6.13. Accuracy of the sensitivity function graph of fault BBF

6.6.2.2 Sensitivity analysis of fault DCP

In the sensitivity analysis of fault DCP, it is given that the interesting parameter DCP =1, evidence CF = 1, and parameter under study is LC, and in objective 1, the parameter probability (x) is sought. The maximum value of M_1 is 1.484 at $x = 0$ reveals that it has most effect on the parameter under given evidence. The relationship between parameter probability (x) and $f(x)$ is depicted in Figure 6.14. In objective 2, the parameter is sought for given the interesting probability, DCP = 1 and evidence, LC = 1 that maximizes M_2

value is 0.142 under the parameter CF = 1 | (BLB = 1, LC = 1, IC = 1, OL = 1, OH = 0) and the relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.14. In objective 3 for the given interesting parameter DCP = 1, evidence, CF = 1, and the new observed evidence LC = 1, parameter that maximizes M_2 value is 0.230 under the parameter VF = 1 | (LC = 1, IC = 0, OH = 0). The relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.14 reveals that the most effective parameter having new evidence LC = 1 is observed. Hence the parameter becomes more effective when new evidence is observed. The most effective parameter in the interesting node DCP has been validated using sensitivity analysis, as shown in Figure 6.15. The maximum M_2 value was calculated as 0.194 and the error was 15.65%.

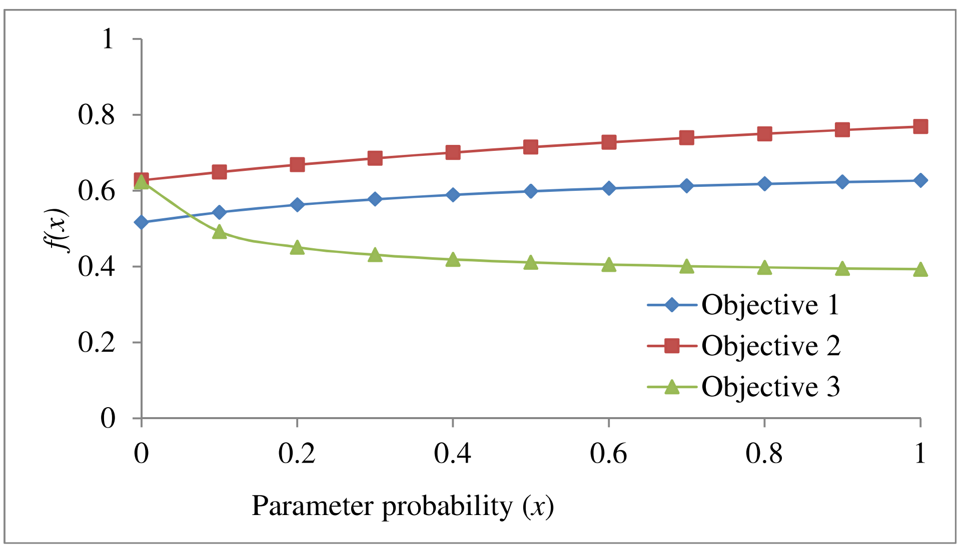


Figure 6.14. Sensitivity function graph of fault DCP for various objectives

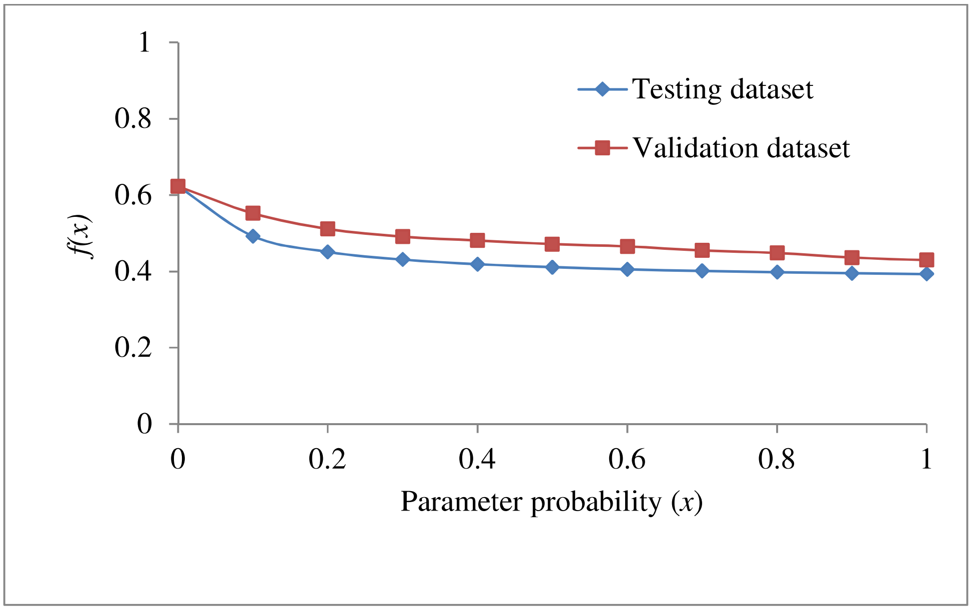


Figure 6.15 Accuracy of the sensitivity function graph of fault DCP

6.6.2.3 Sensitivity analysis of fault SBP

The sensitivity analysis of fault SBP, it is given that the interesting probability SBP =1, evidence VUS = 1, and parameter under study is DC, and in objective 1, the parameter probability (x) is sought. The maximum value of M_1 is calculated as 1.312 at x = 0 means that it has most effect on the parameter under given evidence. The relationship between parameter probability (x) and $f(x)$ is depicted in Figure 6.16. In objective 2, for the given interesting probability SBP =1, and evidence DC = 1, that maximizes M_2 value is 0.166 under the parameter (x), VUS = 1 | (BLB = 0, IC = 1, OL = 1, OH = 1, DC = 1), and the relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.16. In objective 3, for given interesting probability SBP =1, evidence, VUS = 1 and new evidence DC = 1, that maximizes M_2 value is 0.223 under the parameter (x), LL = 1 | (BLB = 0, OH = 1, DC = 1). The relationship between parameter probability (x) and $f(x)$ is depicted in Figure 6.16. Hence the parameter is less effective when new evidence is observed. The most effective parameter under interesting node SBP has been validated using sensitivity analysis as depicted in Figure 6.17. The maximum M_2 value was calculated as 0.213 and error was 4.48%.

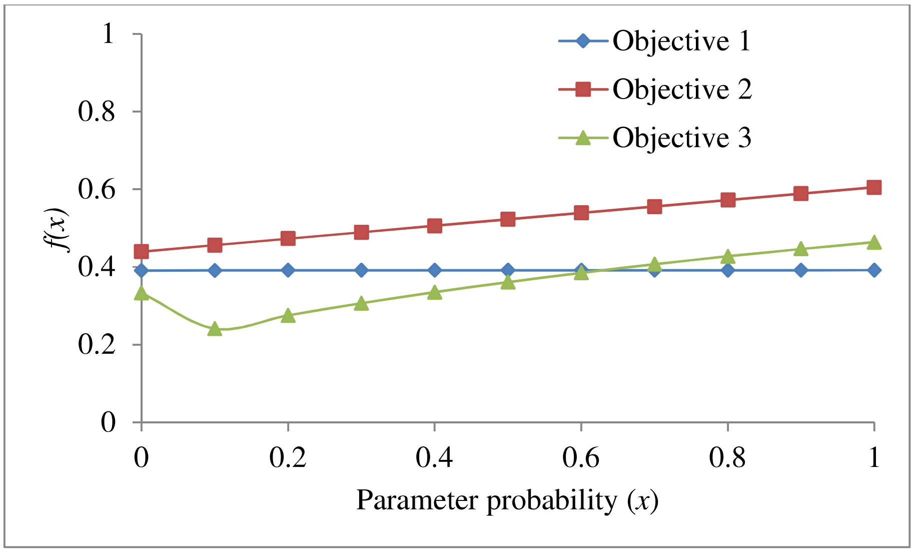


Figure 6.16. Sensitivity function graph of fault SBP for various objectives

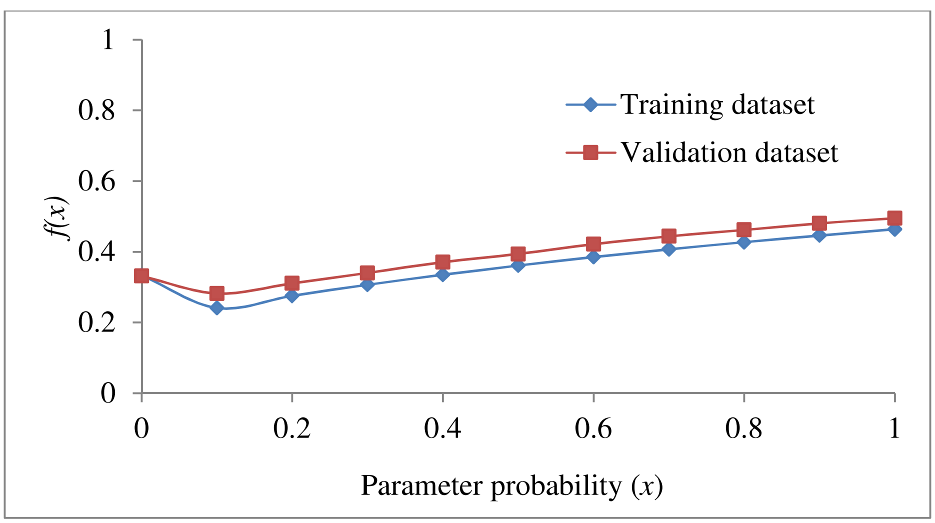


Figure 6.17. Accuracy of the sensitivity function graph of fault SBP

6.6.2.4 Sensitivity analysis of fault ID

The sensitivity analysis of fault ID, it is given that the interesting probability ID =1, evidence CF = 1, and parameter under study is OH, and in objective 1, the parameter probability (x) is sought. The maximum value of M_1 is obtained as 1.78 at $x = 0$ reveals that it has more effect on the parameter under given evidence. The relationship between parameter probability x and $f(x)$ is shown in Figure 6.18. In objective 2 for given interesting probability ID =1, and evidence OH = 1, parameter that maximizes M_2 value as 0.482 under the parameter(x) CF = 1 | (BLB = 0, LC = 1, IC = 1, OL = 0, OH = 1), and the relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.18. In objective 3 for the given interesting probability ID =1, evidence CF = 1, and if a new evidence OH = 1 is observed, then the parameter that maximizes M_2 value is 0.079 under the parameter TF = 1 | BLB = 0, IC = 0, OL = 0, OH = 1, DC = 0). The relationship between parameter probability (x) and $f(x)$ is shown in Figure 6.18. Therefore, the parameter is more effective when new evidence is observed. The most effective parameter under interesting node ID has been validated using validation dataset, as shown in Figure 6.19. The maximum M_2 value was 0.088 and error was 11.39%.

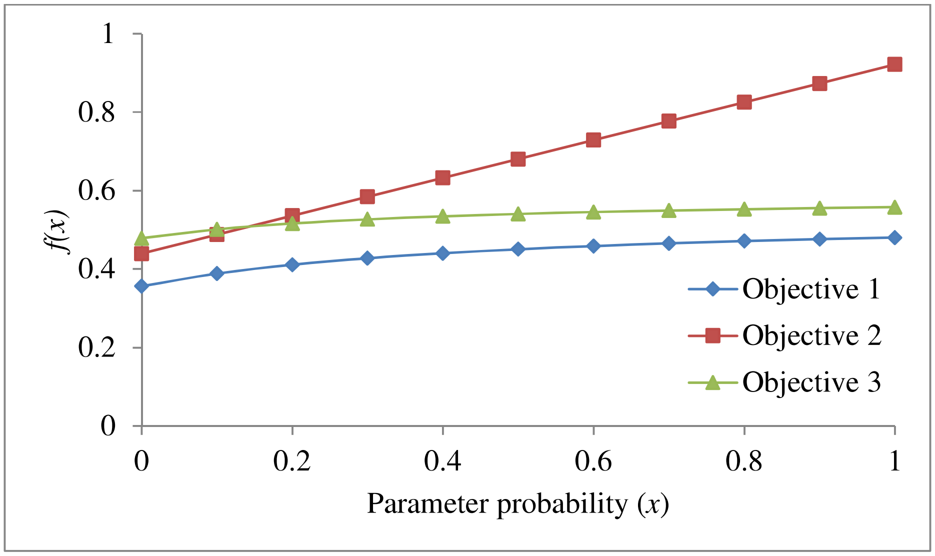


Figure 6.18. Sensitivity function graph of fault ID for various objectives

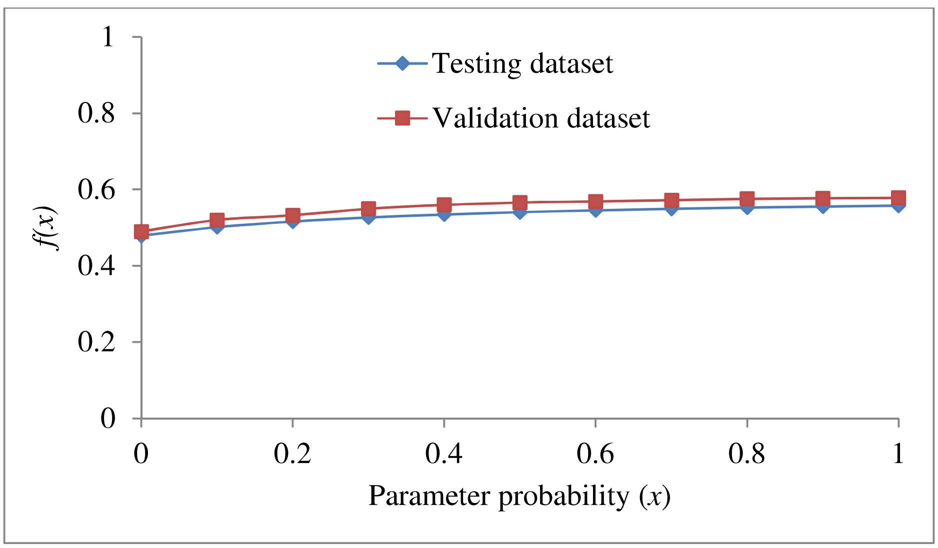


Figure 6.19 Accuracy of the sensitivity function graph of fault ID

Similarly, the fault types for all possible combinations of cases for a given set of evidence can be identified that will help to develop the structure of maintenance strategy. After finding the parameter value that maximizes the posterior probability of the interesting node, the value of the parameter can be controlled or adjusted by adjusting operations of the dragline during maintenance.

6.7 Summary

The fault analysis of drag system was carried out using historical cause, symptom, and fault data. The data was generated continuously for every 15 min from the sensor feedback, maintenance worksheet, and visual inspection. The collected data are categorized either '0' or '1' based on the threshold limit value. Six cause nodes, six symptom nodes, and four fault nodes are identified. The BN model is constructed to make the causal relationship between cause, symptom and fault by estimation of CPT.

The conflict analysis is used to measure the conflict among the observed set of evidences as well as to validate the BN model. The fault inference of BN model is used to make reasoning to categorize fault such as catastrophic fault, degraded fault or intermittent fault based on the observed evidence. The experts' opinion helped to decide the degree of detection limit of fault such as α, β, and ε, which is further validated for some specific fault cases. After identification of fault types, the decision support system can help to make a decision for either to continue to operate the dragline or to discontinue the operation to prevent a major failure. The maintenance type, e.g. urgent maintenance or maintenance is done during schedule inspection interval or during P-F interval of fault is designated based on the fault type.

Finally, the sensitivity analysis of the BN model is used to identify the most influencing parameter that is responsible for the occurrence of fault for given observed evidence and new evidence identified during maintenance. The BN model is also validated using three axioms based sensitivity analysis. The accuracy of BN model is calculated in terms of percentage of error and the model is also validated using validation dataset for the most effective parameter. Finally, case-based reasoning is proposed to be used in the decision support system to prepare the suitable CBM policy.