MULTI-OBJECTIVE SENSOR ARRAY OPTIMIZATION UNDER UNCERTAINTY

by

NAFI AHMED

A thesis
Submitted to the
Department of Industrial and Production Engineering,
Bangladesh University of Engineering and Technology,
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

in

Industrial and Production Engineering

Department of Industrial and Production Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

December, 2020

CERTIFICATE OF APPROVAL

The thesis titled "**MULTI-OBJECTIVE SENSOR ARRAY OPTIMIZATION UNDER UNCERTAINTY**" submitted by Nafi Ahmed, Student No. 0417082008, Session: April 2017, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Science in Industrial and Production Engineering on December 28, 2020.

Board of Examiners

Dr. AKM Kais Bin Zaman
Professor
Department of IPE, BUET, Dhaka — Chairman (Supervisor)

Dr. Nikhil Ranjan Dhar
Professor and Head of the Department
Department of IPE, BUET, Dhaka — Member (Ex-officio)

Dr. Ferdous Sarwar
Associate Professor
Department of IPE, BUET, Dhaka — Member

Dr. Syed Mithun Ali
Associate Professor
Department of IPE, BUET, Dhaka — Member

28/12/2020

Dr. Md. Forkan Uddin
Professor
Department of EEE, BUET, Dhaka — Member (External)

CANDIDATE’S DECLARATION

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Nafi Ahmed
Student ID: 0417082008

ABSTRACT

In a sensor system, the selection of appropriate sensors is very important to obtain a better classification performance. An optimized set of sensors is necessary for accurate analysis of different analytes. Adding many sensors to sensing systems does not improve the accuracy of the classification. On the contrary, the noise generated from the redundant sensors negatively affect the accuracy of the classification. In this research, robust and reliability-based multi-objective sensor array optimization models are proposed to optimize the sensor arrays under uncertainty. Both selectivity and diversity criteria have been considered for constructing the objectives functions. A sensor system prototype capable of detecting analytes like smokes and volatile organic compounds has been designed and used to demonstrate the proposed model. A statistical criterion, general resolution factor (GRF) and Principal Component Analysis (PCA) are used to evaluate the optimization results. The experimental results indicate that the proposed methods can successfully identify the Pareto optimal solutions and an optimized set of sensor array, providing improved input quality for the pattern recognition.

ACKNOWLEDGEMENTS

By the grace of the most benevolent and Almighty Allah, the thesis titled "MULTI-OBJECTIVE SENSOR ARRAY OPTIMIZATION UNDER UNCERTAINTY" has been completed.

The author of this thesis would like to express his sincere gratitude to his thesis supervisor Dr. AKM Kais Bin Zaman, Professor, Department of Industrial & Production Engineering for his whole-hearted supervision. His understandings, encouragements, guidance and instructions throughout the progress of the thesis and report writing have provided a good basis for this research. His inputs during the development of the idea in this thesis have contributed substantially to the completion of the author's work.

The author would also like to convey his sincere gratitude to Dr. Nikhil Ranjan Dhar, Professor and Head, Department of IPE, BUET, Dr. Ferdous Sarwar, Associate Professor, Department of IPE, BUET, Dr. Syed Mithun Ali, Associate Professor, Department of IPE, BUET, and Dr. Md. Forkan Uddin, Professor, Department of EEE, BUET, for their constructive remarks and kind evaluation of this study. The author is also indebted to Nazmul Hasan, Lecturer, Department of IPE, BUET for his valuable suggestions in performing this thesis work.

The author would like to express thanks and appreciation to all well-wishers who inspired him to continue this work. He is grateful to his family members for their support and love, which was an important source of inspiration. Lastly, the author wants to convey his gratitude to all those who have supported him in any respect during the research work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MQ	:	Message Queue
DO	:	Deterministic Optimization
RDO	:	Robustness-based Design Optimization
RBDO		Reliability-based Design Optimization
FCBF	:	Fast Correlation-Based Filter
GRF	:	General Resolution Factor
PCA	:	Principal Component Analysis
GA	:	Genetic Algorithm
VOC	:	Volatile Organic Compounds
ADC	:	Analog to Digital Conversion
FORM	:	First Order Reliability Method
RIA	:	Reliability Index Approach
PMA	:	Performance Measure Approach
PDF	:	Probability Density Function
CDF	:	Cumulative Distribution Function
MPP	:	Most Probable Point
CO	:	Carbon monoxide
LPG	:	Liquefied Petroleum Gas

NOMENCLATURE

C_j	=	Concentration of vapor in Parts Per Million (PPM)
d	=	Vector of deterministic design variables ($\boldsymbol{x}_d$) as well as the mean values of the random design variables ($\boldsymbol{x}_r$), i.e., $\boldsymbol{d}$=[$\boldsymbol{x}_d$ $\boldsymbol{\mu}_{xr}$]
$E(g_i)$	=	Expected value or mean of the ith inequality constraint
f	=	Objective function
g	=	Inequality constraint
h	=	Equality constraint
$H(\tilde{S})$	=	Entropy of sensors
H_{max}	=	Maximum entropy for a sensor array
i	=	Index of vapor
j	=	Index of sensor
K	=	Subset of sensors
$\boldsymbol{lb}$	=	Vector of lower bounds of design variables
$\boldsymbol{LB}$	=	Vector of lower bounds of inequality constraints
M	=	Total number of sensors
N	=	Number of Vapors
$nddv$	=	Numbers of the deterministic design variables
$nrdv$	=	Numbers of the random design variables
$\boldsymbol{p}$	=	Vector of random design parameters
Pfj	=	Failure probability for jth inequality constraint
$Pfj^{\text{allowable}}$	=	Allowable (maximum) failure probability for jth inequality constraint
$\boldsymbol{r}$	=	Vector of all the random variables, i.e., $\boldsymbol{r}$=[$\boldsymbol{x}_r$ $\boldsymbol{p}$]
R_{ij}	=	Response of sensor i to the analyte j
Rs	=	Sensor resistance at various concentrations
RL	=	Sensor load resistance
S_{ij}	=	Sensitivities of j sensor towards i vapor

NOMENCLATURE (CONTINUED)

$\boldsymbol{u}$	=	Standard normal space vector corresponding to the vector of random variables, $\boldsymbol{r}$
$\lVert\boldsymbol{u}\rVert$	=	Euclidean norm of vector $\boldsymbol{u}$
$\boldsymbol{u}*.$	=	Most probable point (MPP) of failure
$\boldsymbol{ub}$	=	Vector of upper bounds of design variables
$\boldsymbol{UB}$	=	Vector of upper bounds of inequality constraints
V_c	=	Circuit voltage
V_{RL}	=	Voltage sensor in the sample space
w	=	Weighting coefficient
$\boldsymbol{x}$	=	Vector of design variables
$\boldsymbol{x}_d$	=	Vector of deterministic design variables
$\boldsymbol{x}_r$	=	Vector of random design variables
β	=	Reliability index
μf	=	Mean value of the objective function
$\boldsymbol{\mu}_p$	=	Vector of mean values of random design parameters
$\boldsymbol{\mu}_{xr}$	=	Vector of mean values of the random design variables
σ_f	=	Standard deviation of the objective function
$\sigma(\boldsymbol{x}_r)$	=	Vector of standard deviations of the random design variables
Φ	=	Standard Gaussian cumulative distribution function (CDF)
z	=	Vector of non-design input random variables

CHAPTER 1
INTRODUCTION

1.1 Background of the Study

Sensor array systems are efficient and effective tools to detect chemical or biological constituents, e.g., volatile organic compounds (VOCs), in industrial control or in monitoring process. Sensing system has been successfully applied in different industries like food processing, pharmaceuticals, bio-medical, textile and leather manufacturing, etc. [1]. Its Industrial and Production Engineering applications include system optimization, process control, data analysis, industrial health & safety analysis, and uncertainty & reliability analysis, etc. [2]. Generally, a sensing system usually consists of three main components: an array of sensors, signal processing, and pattern recognition [3]. Sensor array optimization is the most important, because the sensor array responses directly determine the input quality for the pattern recognition [1].

The primary purpose of sensor array optimization is to select appropriate sensors with specific attributes [4]. Generally, the optimized set of sensors provides less entropy in the sensor system [3]. This sensor selection is a special type of feature selection process [5]. Different statistical methods and stochastic algorithms have been used in sensor array optimization [6]. With the growing complexity of optimization problems, stochastic techniques are becoming more and more popular. These methods are suitable for sensor array optimization because they can be easily adapted to identify the optimal variable set based on certain selection criteria [7, 8].

In sensor array optimization, the sensitivities of different sensors are considered [3]. These sensitivities of the sensors may vary. Therefore, the uncertainty in the input variables is essential to be taken into account. Robustness-based optimization [9] and reliability-based optimization [10] are two approaches that can deal with optimization under uncertainty. However, much of the work has not been done yet on the robust and reliability-based sensor system optimization processes [11]. Therefore, development of robust and reliability-based multi-objective sensor array optimization model is still an open problem and thereby yields the scope of the proposed thesis.

1.2 Objectives with Specific Aims

The specific objectives of this research are:

(i) To develop robust and reliability-based multi-objective sensor array optimization formulation for minimizing entropy based on both selectivity (sensor's response to the target analyte) and diversity (sensor's response to the rest of the analytes) criteria.

(ii) To design a sensor system capable of detecting smokes and different volatile organic compounds (VOCs).

(iii) To select suitable sensors for the designed sensor system based on optimization results and remove redundant sensor elements to minimize the noise and distortions of the signals.

1.3 Possible Outcomes

The current research has developed and demonstrated robust and reliability-based approaches for multi-objective sensor array optimization. The framework is expected to contribute in various domains in industrial and production engineering, for example, system optimization, process control, industrial health and safety, uncertainty analysis, etc.

1.4 Outline of Methodologies Used

The methodologies used in this research is outlined below:

(i) A deterministic multi-objective optimization problem has been formulated with the objectives being minimization of entropy based on both selectivity and diversity criteria using Shannon's information entropy theory and concept of information gain.

(ii) A filter type feature selection process called Fast Correlation-Based Filter (FCBF) algorithm has been used for selecting the members of the sensor arrays.

(iii) The input variables, parameters, functional relationships and uncertainty associated with the input variables have been identified using existing literature.

(iv) A robustness-based and a reliability-based multi-objective sensor array optimization have been formulated with the objectives being minimization of entropy based on both selectivity and diversity criteria of the sensors.

(v) A system of sensors capable of detecting smokes and other VOCs (Volatile Organic Compounds) has been designed using MQ (Message Queue) gas sensors and Arduino microcontroller.

(vi) Data have been collected from the sensors of the designed sensor system in (v).

(vii) The robust and reliability-based multi-objective optimization methodology developed in step (iv) have been used for selecting suitable sensors for the designed sensor system.

(viii) Deterministic optimization, robustness-based formulation and reliability-based formulation developed in step (i) and step (iv) have been solved using MATLAB.

(ix) A statistical evaluation method, general resolution factor (GRF) and Principal Component Analysis (PCA) have been used to ensure the effectiveness of the optimization methods. GRF and PCA has been used verify the inherent quality of the input spaces for pattern recognition.

1.5 Contributions of the Present Study

This thesis contributes to developing robust and reliability-based multi-objective sensor array optimization formulations to minimize entropy based on both selectivity and diversity. The optimization model considers the uncertainties present in the input variables which is more appropriate than previously developed formulations in the literature. The consideration of uncertainty in design variables is a new addition to the sensor array optimization. The incorporation of uncertainty has enabled the model to be more robust and reliable in practical scenarios.

The proposed model can identify a set of optimal sensors and remove redundant sensor elements to minimize the noise. The noise created by the redundant sensors are responsible for the poor performance of the sensor system. Therefore, removing excess sensors from the senor system ensures the improvement in solution quality and a better classification performance in the sensor systems. Moreover, industries will be benefitted economically by removing excess sensors without sacrificing quality.

1.6 Organization of the Thesis

The thesis report is organized in the following manner:

Chapter 1 contains the necessary background of the thesis, specifically defined objectives, possible outcomes, a summary of the developed methodology, and the contributions of this study in the field of sensor array optimization. Chapter 2 contains a brief review of all the relevant literature. Chapter 3 provides the necessary theoretical background regarding optimization under uncertainty. The detail of the proposed methodology is described in Chapter 4. In Chapter 5, the proposed methodology is illustrated using a sensor system prototype and the results are discussed and evaluated. Finally, in Chapter 6, the thesis is concluded with recommendations for future work.

CHAPTER 2
LITERATURE REVIEW

2.1 Review on Sensor Array Optimization:

Sensor array systems are highly sensitive sensing systems for detecting and quantifying toxic industrial chemicals with even very low concentration levels [12]. As mention earlier in Chapter 1, a sensing system mostly consists of three major components: a sensor array, signal processing, and pattern recognition. The sensor array is the most critical component among them since it responds to a wide range of chemical mixtures. Its response affects the input quality of the pattern recognition and classification part [3]. In other words, the classification accuracy is affected by the combination of the sensor array and is the most important performance measure of the sensing systems [1]. Adding many sensors to sensing systems does not boost the accuracy of the classification; quite the opposite, due to the collinearity that has a negative impact on the accuracy of the classification, some sensors may contain redundant or irrelevant information [1, 13]. Therefore, the sensor array optimization is the most significant step of sensing system optimization.

Many studies have explored different methods for sensor array optimization to date. Lin and Suslick [14] have developed a colorimetric sensor array to detect tri-acetone triperoxide (TATP), one of the most dangerous primary explosives and is very difficult to detect. The traditional techniques are either expensive, either require a long sample preparation process or cannot detect TATP in the gas phase. However, the colorimetric sensor array is capable to detect TATP among 18 interferences (such as humidity, perfume, lotion, vinegar, laundry supplies, volatile organic compounds, etc.) [14]. Another study detected 20 toxic industrial chemicals (TICs) with 36 sensors at their PELs (permissible exposure limits) by developing a colorimetric sensor [15]. Commonly in food quality monitoring, Carbon Dioxide (CO_2) is the main indicator of food quality [16, 17]. In addition, air pollutants, e.g., Hydrogen Sulfide (H_2S) and Ammonia (NH_3) as results of the protein decomposition, are an indicator of bacterial metabolites [18, 19]. The challenge is how to find the best-suited combination of sensors from several sensors with overlapping selectivity.

Stochastic techniques are gaining popularity to identify the best combination of sensors [1]. Xu et al. [3] employed genetics algorithm (GA) to optimize the sensor arrays and to remove the redundant sensors. Cluster analysis was also employed in their study to find the amount of sensor groups. They have used a statistical measure: General Resolution Factor (GRF) and Principal Component Analysis (PCA) to validate their optimization results. Another sensor array optimization analysis used 15 gas sensors for wound infection detection [20]. It combined genetic algorithms (GA) and quantum-behaved particle swarm optimization (QPSO) to create synchronous optimization between detectors and classifiers. This study proposed the weight of sensor as a degree of importance in classification. The traditional method of weighting offers 0 and 1 to mark the contribution of the sensors. If the weight of the sensor is 0, this means that no contribution is made. Conversely, 1 implies the full contribution of the sensor. Their research used real numbers for weighting coefficients instead of using conventional weighting approaches. It is called the Importance Factor. The optimized sensor array based on the Importance Factor approach had an accuracy improvement equal to 7.5 percent with using SVM as a classifier. In addition, some studies have focused on problems with the optimization of the sensor array using function selection techniques, such as the use of the statistical and heuristic models like the Rough Set-Based approach [21], the Neural Network [22], the combination of function selection methods, etc. [23].

Many researches employed the feature selection techniques in sensor array optimization. The main advantages of feature selection incudes increasing the speed of algorithm, minimizing the computation resources (e.g., memory, storage, and processor), obtaining a higher classification accuracy, and simplifying the data visualization [24]. To find the best subset of relevant features, feature selection techniques such as Filter, Wrapper, and Embedded are commonly used. In addition, because of its robustness, the filter technique is a promising approach to solving the over-fitting problem in classification [25, 26]. The uncertainty and annoyance caused by the over fitting hampers the classification performance. The key idea of the filter technique is generally the removal of redundant characteristics by quantifying the association between the characteristics.

Saha et al. [27] have analyzed a sensor selection process by using three feature selection methods to select the most important sensors. They used multi-class support vector machine (SVM) to identify the observations, which are *t*-statistics, minimum redundancy maximum relevance techniques, and Fisher's criterion. Sathiskumar et al. [28] have introduced *k*-means, fuzzy *c*-means, and rough *k*-means clustering methods for the gas sensor array drift data collection. There are two different factors in optimal sensor systems, which are selectivity and diversity. In sensor array systems, highly selective sensors are desired as they produce the most important information among the large range of mixtures for each chemical analyte. Therefore, if each sensor has a different sensitivity to each analyte [3], the sensor array system as a whole will generate specific response patterns for different analytes. This will increase the variety of the chemical analytes whose reaction patterns are successfully detected by the array of sensors. Therefore, the sensors which improve both the selectivity and diversity of the sensor array systems are selected in this study. However, much of the work has not been done considering the uncertainties in the input variable for sensor array optimization. However, in practical scenario, these uncertainties need to be considered for better results. Therefore, robustness and reliability-based optimization, which are capable of dealing with optimization under uncertainty, would be a good addition to the literature.

2.2 Review on Optimization Under Uncertainty:

The current research is intended to develop a methodology of multi-objective sensor array optimization under uncertainty. This uncertainty prevails in the natural characteristics of a design and observed data and this is often ignored in the deterministic optimization. Optimization under uncertainty has gained increasing attention in the last few decades due to this reason. There are now an extensive volume of methods and applications available for optimization under uncertainty. Robustness-based design [29, 30, 31, 32] and reliability-based design [32, 33, 34, 35] are two prominent fields of optimization that consider the design parameter uncertainty. Robustness is the performance criteria for a system to operate continuously for a wide range of operational conditions and will be failed outside the conditions [36]. Taguchi developed the concept of robust design and proposed a method where the product performance or the output remains insensitive to the variation in design variables in the manufacturing process [37]. The variation in the design variables was designated as noise, which could be created from various factors in the manufacturing process.

As all of the engineering models are becoming more and more complex day by day, applying statistical design tools in Taguchi's method is not well enough to calculate optimal feasible solution for multiple measurements of performance and design constraints [38]. It was possible to achieve robustness in both performance outputs and design constraints due to the implementation of nonlinear programming in a robust design. [30]. There are normally some significant variables or elements in the system which cause uncertainty in the system. Uncertainty may arise from two sources, which are aleatory and epistemic [39]. Aleatory uncertainty cannot be reduced due to including natural phenomena that exhibit natural variation like operating condition, material properties, geometric tolerances, etc. on the other hand, epistemic uncertainty arises from a lack of knowledge about the system or due to approximations in the system behavior models, or due to limited or subjective data. However, this study only considers the aleatory uncertainty. Epistemic uncertainty is out of the scope of this research.

2.2.1 Review on robustness-based design optimization

Zaman et al. [40] described four key elements of robustness-based design optimization : (1) maintaining robustness in the objective function, which is called objective robustness; (2) achieving robustness in the constraints, which is called feasibility robustness; (3) estimating the mean and the variance of the performance function; and (4) multi-objective optimization. In robustness-based optimization, objective robustness can be achieved by simultaneously optimizing the mean and minimizing variance of the objective function. Two types of robustness measurement processes are popular: one is based on the variance [30,31], and the other is based on the percentile difference [41]. On the other hand, feasibility robustness can be defined as the robustness in the constraints. Robustness in the constraints indicates satisfying the constraints of the design in the presence of uncertainty. Du and Chen [30] classified the methods of maintaining feasibility robustness into two categories: (i) methods based on stochastic and statistical analysis, e.g., a probabilistic feasibility formulation [42], and a moment matching formulation [29] and (ii) methods that do not use probabilistic and statistical analysis. Several methods have been developed which are not depending on probabilistic and statistical analysis like worst-case analysis [29], corner space evaluation [43], and manufacturing variation patterns (MVP) [44]. The feasible region reduction method [45] is general in application and normality assumption is not necessary. This is a tolerance design method, where the width of the feasible space in each direction is reduced by the amount of $k\sigma$, where k is a user defined constant and σ is the standard deviation of the performance function.

The mean and the standard deviation or variance can be estimated through several methods. There are some existing methods for estimating the mean and standard deviation of performance function, which can be classified into three major categories: (1) Taylor series expansion methods, (2) Sampling based methods, and (3) Point estimate methods [32]. The Taylor series expansion method [30, 45, 46] is quite appreciable to measure the mean and variance of nonlinear performance function. However, the approximation may result in huge error if the variances of the random variables are large [41]. Information on distribution of the random variables is required to estimate mean and variance by the sampling-based methods, which has made it expensive. Point estimate method [47] can be used to ease the computation of the derivatives required in the Taylor series expansion. Several types of point estimate methods [48] can be applied to estimate the mean and variance. Dimension reduction method (DRM) [49] is developed which overcomes the troubles associated with Taylor series expansion and sampling method. Multi-objective optimization can be achieved by optimizing the mean of the objective function and minimizing its variation [50]. Weighted sum approach has widespread application in multi-objective optimization for robust design problems [51]. Other optimization methods called ε-constrained method [52], goal programming [51], compromise programming (CP) [53] and physical programming [54] can be enlisted. There are different advantages and shortcomings of all those methods.

2.2.2 Review on reliability-based design optimization

Reliability-based design optimization (RBDO) uses numerical optimization algorithms to obtain optimal design with reliability [55]. The reliability optimization is usually performed to ensure a safety limit or target reliability for deterministic data. However, without considering the uncertainty may lead to system failure. Therefore, it is necessary to include uncertainty in the design constraints for having a reliable design [56]. The uncertainties can be modeled and represented using the probability theory. The reliability optimization may require to optimize a single or multi-objective function while satisfying the reliability constraints. The reliability constraints consider the probability of failure (P_f), which is related to the system's failure mode or design [57]. Different simulation methods with high computational abilities like Monte Carlo Simulation (MCS), importance sampling, etc., can be used to get a reliable solution. In the stochastic design, it is necessary to estimate the uncertainty of the random variables. The probability-based approach can have contributions to represent the uncertainty. After assessing the uncertainty, the reliability based design optimization are employed to obtain an optimal and reliable solution [55].

The key challenge of an RBDO problem is the assessment of a design's likelihood of failure as it takes a considerable effort to compute. Conventionally, under probabilistic constraints, RBDO problems are formulated as a stochastic optimization problem, where the solution is obtained by including the failure probability evaluation in the main optimization loop. Usually, this method leads to a nested optimization problem, referred to as the computationally extensive double loop approach [58]. RBDO based on First Order Reliability

Method (FORM) is a nested double loop approach that requires an optimization loop as well as a reliability loop, i.e. two optimizations are coupled. The well-known methods based on FORM technique are the Reliability Index Approach (RIA) [59] and the Performance Measure Approach (PMA) [60]. To reduce the computational cost, two approaches have been proposed. The first one is to separate the reliability assessment from the optimization loop and convert the RBDO problem into the sequences of deterministic optimization and reliability assessment cycles. This approach, namely the decoupled approach, is the key idea of the Sequential Optimization and Reliability Assessment (SORA) method [61]. The second approach, known as the single loop approach, consists of converting the probabilistic constraints into deterministic ones. Thus, the RBDO problem becomes a deterministic optimization problem. There are two main methods that use this approach: the Single Loop Approach (SLA) proposed by Liang et al. [62,63] and the Reliable Design Space (RDS) method proposed by Shan et al. [64]. However, Reliability Index Approach (RIA) approach is still more popular in the literature due to its simplicity and efficacy [59]. In this thesis, RBDO based on FORM method that follows the Reliability Index Approach (RIA) approach has been used.

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Sensor Array Optimization

Sensor array optimization is performed by selecting suitable descriptors, i.e., sensors, each with a particular attribute. Sensor selection is a special type of feature selection for input. In general, the collection of features will find a subset of appropriate characteristics with a minimum size suitable for the target [65]. The resulting input with small dimensionality can reduce system complexity through feature selection, while retaining or improving prediction accuracy [66]. The general procedure of feature selection is composed of four parts:

1. Generation of candidate subset
2. Evaluation of subset
3. Finding a stopping criterion
4. Validation of subset

There are two basic types of methods for feature selection: filter and wrapper [67, 68]. In order to test subsets, the wrapper approach uses specific classifiers, so the prediction accuracy for that specific classifier is high after selection. Wrapper strategies, however, usually have high computational costs and limited generalization. The method proposed in this thesis belongs to the filter type, which is suitable for general classification applications. During optimization, the efficiency of the different sensor arrays is evaluated by information measure, and the final selection result is accessed by independent validation through both statistical and visual inspection. Till now, only a few studies have addressed multi-objective optimization model on sensor selection, and most of them used single-objective optimization methods [3]. There appears to be only one ultimate goal in sensor array optimization: find a subset of sensors providing optimal input for classification. However, this aim can be best accomplished by simultaneously optimizing multiple targets that have a direct effect on the quality of the input. In other words, by simultaneously following multiple requirements that can guarantee high input feature space efficiency, the sensor array can be optimized. In application, the majority of optimization problems naturally require multiple objectives or goals. These goals are almost equally significant, and they are probably in conflict with each other. Generally, there is no common optimal solution for all objectives. Multi-objective, or multi-criterion optimization deals with this situation by locating a set of solutions with trade-offs among objectives. The general procedure of a multi-objective optimization includes two basic steps [69]:

Step 1: Finding a Pareto optimal set consisting of solutions with compromise concerning different objectives.

Step 2: Evaluating and choosing among those solutions based on higher-level information.

After Step 1, a set of solutions is found for examining the trade-offs with respect to different objectives. In this way, the drawbacks of converting multi-objective into single-objective can be avoided. In this thesis, the weighted sum approach has been used to solve the multi-objective problem. Then, in Step 2, some qualitative and subjective data can be used by decision makers to access and choose the final solution. Multi-objective optimizations have recently been used successfully in a wide variety of applications, such as transportation, production planning, scheduling, and forecasting, etc. [70]. For sensor array optimization, single-objective methods might fail to locate the true optimal sensor combinations, because they do not consider all significant factors. Selectivity and diversity are two such important factors for a sensor array. A sensor with high selectivity is needed for a wide range of vapors because it can provide the most important information about each vapor. A separate response pattern for each vapor should be provided by the array as a whole in the design of the sensor array. The pattern of response consists of responses from all sensors. If every sensor has different sensitivities with respect to each analyte, a sensor array would have distinct response profiles to different analytes. Based on their response patterns, the diversity allows the sensor array system to successfully recognize various vapors [1].

3.2 Design Optimization

Optimization is a methodology that uses a mathematical description of a system to find the best possible solution based on the system's characteristic criterion, subjected to recognized constraints. Mathematical description of a system refers to an abstraction of a real system using mathematical expressions of relevant natural laws, physical properties, accumulated empirical evidence, and geometric features. A real system is difficult to analyze since a system existing in a real environment can undergo very complex situations. Thus, a mathematical description or model is essential to represent a system to increase the understanding of how a system works. Mathematical model representing a system generally includes the following elements—

1. Variables: These are entities that can assume different values within acceptable ranges and thus affect the system condition. For example— the sensitivities of a sensor.
2. Parameters: These are entities that are given value set by the modeler to define a specific condition for the system. For example— maximum entropy value in a sensor system.
3. Constants: These are entities fixed by natural laws, physical properties, and geometric features. For example— Shannon's constant in entropy information theory.
4. Mathematical Relations: These are mathematical expressions that include equalities, and inequalities involving system variables, parameters, and constants. These relations are used to describe the system's objective function, and the constraint functions represent the limitations imposed by its environment.

Optimization requires a characteristic criterion of the system to be selected so that it can be used to compare the available alternatives for finding the best solution. The selected criterion is referred to as 'objective'. The mathematical relation representing the objective is known as 'objective function'. The system under consideration can be subjected to limitations enforced by the natural laws, availability of material characteristics, and geometric compatibility. Thus, there exists a set of requirements that must be satisfied by any acceptable design. The conditional requirements are represented by the mathematical relation known as 'constraint function'. The term 'function' does not necessarily refer to a single mathematical relation; rather, it may be a system of equations. Therefore, the number of constraints and objective functions may be one or more. The steps involved in the formulation of a typical deterministic optimization problem can be summarized as follows:

Step 1: Selecting a set of design variables from the system variables to describe the design alternatives.

Step 2: Selecting an objective function to be minimized or maximized. The objective function is expressed in terms of related design variables, parameters, and constants.

Step 3: Determining a set of constraint functions that must be satisfied by any acceptable design. The constraint functions are expressed in terms of related design variables, parameters, and constants.

Step 4: Determining a set of values for the design variables, which minimize or maximize the objective function and satisfy all the constraint functions simultaneously.

An optimization problem can be formulated using different methodologies according to the nature of the problem. The following categories of optimization are discussed in this thesis—

i. Deterministic optimization
ii. Robustness-based design optimization (RDO)
iii. Reliability-based design optimization (RBDO)
iv. Multi-objective optimization

These four categories of optimization have been discussed in the following sections of this chapter.

3.2.1 Deterministic optimization

The input variables in the designs and processes are considered as fixed quantities in the deterministic optimization formulation without considering any stochastic characteristics or data uncertainty in the variables. The deterministic optimization formulation can be written as:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \tag{3.1}$$

$$s.t.\;\; g_i(\boldsymbol{x}) \leq 0;\text{ for all } i \tag{3.2}$$

$$h_i(\boldsymbol{x}) = 0;\text{ for all } i \tag{3.3}$$

$$\boldsymbol{lb} \leq \boldsymbol{x} \leq \boldsymbol{ub} \tag{3.4}$$

where, $f(\boldsymbol{x})$ is the objective function, $\boldsymbol{x}$ is expressed as a vector for design variables, $g_i(\boldsymbol{x})$ is the ith inequality constraint, $h_i(\boldsymbol{x})$ is the ith equality constraint, and $\boldsymbol{lb}$ and $\boldsymbol{ub}$ are the vectors of lower and upper bounds of the design variables.

3.2.2 Robustness-based design optimization (RDO)

In real situations, the input variables are normally uncertain and solution obtained from the deterministic design optimization could be sensitive to the variations of the input variables. Uncertainty analysis could be an important issue in the robustness-based design optimization. The deterministic design optimization could be used to get an optimal point which might be applied as an initial guess in the robustness-based optimization problem. In practice, the robust design optimization is complex in nature which requires nonlinear optimization. In this thesis, objective robustness is achieved by measuring the variation in the objective function through the variance or standard deviation. Feasibility robustness is achieved by feasible region reduction method. First order Taylor series expansion is used to estimate the mean and variance of the objective function. Weighted sum method is used to trade off the multiple objectives in the performance function of the robust design optimization. The formulation of robustness-based design optimization considering only aleatory uncertainty is [40]:

$$\min_{\boldsymbol{x}} f\ (\mu, \sigma) = w \times \mu_f + v \times \sigma_f \tag{3.5}$$

$$s.t.\ \boldsymbol{LB}+k\sigma\ (g_i\ (\boldsymbol{d},\ \boldsymbol{z})) \leq E\ (g_i\ (\boldsymbol{d},\boldsymbol{z})) \leq \boldsymbol{UB} - k\sigma(g_i(\boldsymbol{d},\ \boldsymbol{z}))\ \text{for}\ all\ i \tag{3.6}$$

$$\boldsymbol{lb_i} + k\sigma\ (\boldsymbol{x}_i) \leq \boldsymbol{d_i} \leq \boldsymbol{ub_i} - k\sigma(\boldsymbol{x}_i) \quad \text{for}\ i = 1, 2, 3\ldots\ldots nrdv \tag{3.7}$$

$$\boldsymbol{lb_i} \leq \boldsymbol{d_i} \leq \boldsymbol{ub_i} \qquad \text{for}\ i = 1, 2, 3\ldots\ldots nddv \tag{3.8}$$

where, μ_f and σ_f are the mean value and standard deviation of the objective function, respectively; $\boldsymbol{d}$ is the vector deterministic variables which can be the mean values of the uncertain variables $\boldsymbol{x}$; *nrdv* is the number of random design variables and *nddv* is the number of deterministic design variables; $\boldsymbol{z}$ is the vector of non-design input random variables. $w \geq 0$ and $v \geq 0$ are the weighting coefficients where $w + v = 1$ that represent the relative importance of the objectives μ_f and σ_f respectively; $E(g_i(\boldsymbol{d},\boldsymbol{z}))$ is the mean and $\sigma(g_i(\boldsymbol{d},\boldsymbol{z}))$ is the standard deviation of the *i*th constraint. $\boldsymbol{LB}$ and $\boldsymbol{UB}$ are the vectors of lower and upper bounds of the constraints g_i's; $\boldsymbol{lb}$ and $\boldsymbol{ub}$ are the vectors of lower and upper bounds of the design variables. k is used here to adjust the robustness of the method against the level of conservatism of the solution. Considering the variations in the design variables, k reduces the feasible region and is related to the probability of constraint satisfaction.

3.2.3 Reliability-based design optimization (RBDO)

In the field of optimization, RBDO is a tool that optimizes objective function(s) ensuring that the reliabilities of the inequality constraints or limit state functions are above an acceptable threshold limit, taking into account the uncertainty in design variables and system parameters. A reliability of 0.99 for an inequality constraint implies that the constraint will not be violated in 99% cases under uncertainty of the input variables. In case of equality constraints, the equality relation must hold under uncertainty of the input variables. Reliability estimation requires an analysis of uncertainty to calculate failure probability that is used in the constraint. A typical RBDO problem can be formulated as:

$$\min_{\boldsymbol{d}} f(\boldsymbol{d}, \boldsymbol{\mu}_p) \tag{3.9}$$

$$s.t. \;\; P_{fj} = P(g_j(\boldsymbol{x}_r, \boldsymbol{p}) \leq 0) \leq P_{fj}^{allowable} \quad \text{for all } j \tag{3.10}$$

$$h_j(\boldsymbol{d}, \boldsymbol{\mu}_p) = 0; \quad \text{for all } j \tag{3.11}$$

$$\boldsymbol{lb} \leq \boldsymbol{d} \leq \boldsymbol{ub}; \tag{3.12}$$

where, f ($\boldsymbol{d}$, $\boldsymbol{\mu}_p$) is the objective function; $\boldsymbol{d}$ is the vector of deterministic design variables ($\boldsymbol{x}_d$) as well as the mean values of the random design variables ($\boldsymbol{x}_r$) i.e. $\boldsymbol{d} = [\boldsymbol{x}_d \;\; \boldsymbol{\mu}_{xr}]$; $\boldsymbol{\mu}_p$ is the vector of mean values of random design parameters ($\boldsymbol{p}$); g_j ($\boldsymbol{x}_r$, $\boldsymbol{p}$) is the jth inequality constraint, and in RBDO literature g_j ($\boldsymbol{x}_r$,$\boldsymbol{p}$) is also referred to as 'limit state function'; P_{fj} and $P_{fj}^{allowable}$ represent the failure probability and allowable (maximum) failure probability for jth inequality constraint, respectively; hj ($\boldsymbol{d}$, $\boldsymbol{\mu}_p$) is the jth equality constraint. The vectors $\boldsymbol{lb}$ and $\boldsymbol{ub}$ represent lower and upper bounds of all the design variables, respectively. Note that, the objective function value is estimated at $\boldsymbol{d}$ and $\boldsymbol{\mu}_p$ in RBDO. Eq. (3.10) is referred to as probabilistic formulation of the inequality constraints or limit state functions and in RBDO literature the functional relation g_j ($\boldsymbol{x}_r$,$\boldsymbol{p}$) is expressed in such a manner that failure or violation of inequality constraint is denoted by g_j ($\boldsymbol{x}_r$,$\boldsymbol{p}$) $\leq$ 0, which is opposite to conventional optimization problem notation.

The failure probability P_{fj} for the jth inequality constraint may be obtained by evaluating the integral in Eq. (3.13), which is the fundamental expression of the structural reliability problem:

$$P_{f_j} = \int_{g_j(\boldsymbol{x}_r, \mathbf{p}) \leq 0} f_r(\boldsymbol{r}) dr \tag{3.13}$$

where, $\boldsymbol{r}$ is the vector of all the random variables (i.e. $\boldsymbol{r} = [\boldsymbol{x}_r \;\; \boldsymbol{p}]$); $f_{\boldsymbol{r}}$ ($\boldsymbol{r}$) is the joint probability density function (PDF) of random vector $\boldsymbol{r}$. Note that, g_j ($\boldsymbol{x}_r$,$\boldsymbol{p}$) can also be rewritten as g_j ($\boldsymbol{r}$). Elements of $\boldsymbol{r}$ are assumed to be statistically independent and normally distributed.

3.2.4 Multi-objective optimization

In many practical applications, a user may want to optimize two or more objective functions simultaneously. These are called multi-objective, multi-criteria or vector optimization problems. Common approaches to multi-objective optimization include the weighted-sum method, ε-constraint method, global criterion method, lexicographic method, etc. In this thesis, the weighted-sum method has been used. The weighted sum approach are discussed briefly in the following subsections

3.2.4.1 Weighted-sum method

This is a method that aggregates all the objective functions by multiplying each of them with a weighting coefficient w_i to form a single function, which is then minimized. The value of the weighting coefficients generally reflects the relative importance of the objectives. The optimization problem is formulated as follows:

$$\min_{\boldsymbol{x}} \sum_{i=0}^{I} w_i f_i(\boldsymbol{x}) \tag{3.14}$$

$$s.t. \;\; g_j(\boldsymbol{x}) \leq 0; \;\; \text{for all } j \tag{3.15}$$

$$h_j(\boldsymbol{x}) = 0; \;\; \text{for all } j \tag{3.16}$$

$$\sum_{i=0}^{I} w_i = 1; \;\; w_i > 0 \tag{3.17}$$

$$\boldsymbol{lb} \leq \boldsymbol{x} \leq \boldsymbol{ub} \tag{3.18}$$

Here, $f_i(\boldsymbol{x})$ is the ith objective function; w_i is the weighting coefficient of ith objective function. The weighting coefficients w can be used in two ways— the user may either set w_i to reflect preferences or systematically alter w_i to yield different Pareto optimal points.

3.2.4.2 ε-constraint method

The ε-constraint approach minimizes the single most important objective function $f_s(\boldsymbol{x})$ with other objective functions treated as constraints. The optimization problem is formulated as follows:

$$\min_{\boldsymbol{x}} f_s(\boldsymbol{x}) \tag{3.19}$$

$$s.t. \;\; g_j(\boldsymbol{x}) \leq 0; \;\; \text{for all } j \tag{3.20}$$

$$h_j(\boldsymbol{x}) = 0; \;\; \text{for all } j \tag{3.21}$$

$$f_i \leq \varepsilon_i; \quad \text{for } i = 1, 2, 3, \ldots I \tag{3.22}$$

Here, $f_s(\boldsymbol{x})$ is the single most important objective function, and other objective functions ($f_i(\boldsymbol{x})$; $i \neq s$) are incorporated as constraints; ε_i is the expected value used as corresponding upper bound for $f_i(\boldsymbol{x})$.

In this research, the weighted-sum method has been used in Deterministic Optimization, Robustness and Reliability-based optimization. The following chapter discusses the proposed methodology of this thesis.

CHAPTER 4

PROPOSED METHODOLOGY

In this study, sensor array optimization is performed based on two criteria: high selectivity and high diversity of sensor array. The flowchart of the proposed method is shown in Figure 4.1. First, cluster analysis is used to get primary grouping information, and degree of similarity between sensors. Through cluster analysis, the maximum size of sensor array can be roughly estimated. Members of each group are selected using Fast Correlation-Based Filter (FCBF) algorithm. First, the deterministic optimization formulation is solved to identify the Pareto optimal solutions. Then, robustness based and reliability based multi-objective optimization is implemented to identify the Pareto sets of solutions.

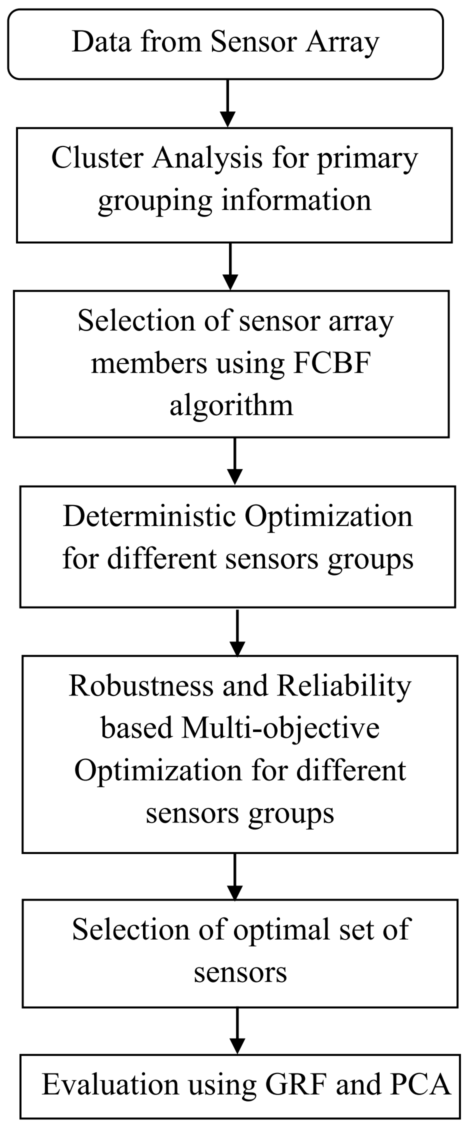

Figure 4.1: Flow chart of the Proposed Method

The solutions obtained from the robustness and reliability-based multi-objective optimization have been used to identify the optimal set of sensors. The performance of the set of sensors are evaluated by a statistical index, General Resolution Factor (GRF). In addition, the classification performance the selected sensors are visually evaluated with the aid of Principal Component Analysis (PCA). The details of each steps of the proposed method have been discussed in the rest of this chapter.

4.1. Cluster analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups [28]. . With respect to some characteristic, observations in each cluster or group are similar to each other; and observations of different groups are different. Usually, similarity is decided by comparing the distances in some space: the smaller the distances between observations, the more similar they are. The most commonly used distance metric is the Euclidean distance. There are two types of cluster analysis:

i. Hierarchical
ii. Non-hierarchical

This study employs the hierarchical clustering to find the number of clusters based on data similarity. The number of clusters is needed to determine the maximum number of features in the feature selection algorithm. Sensors are grouped into several subarrays based on their selectivity. The number of clusters is associated with the optimal amount of gas sensors in the sensor array [71]. To reflect distances between clusters, several methods for developing similarity matrix can be chosen: centroid, single-linkage, complete linkage, average-linkage and Ward's method, etc. [72]. Centroid clustering with Euclidean distance is used in this thesis for its better handling of data. In this research, cluster analysis is employed to preliminarily analyze the approximate groups among all sensor candidates.

4.2. Deterministic Model Formulation

First, the deterministic multi-objective model is developed using the Shannon's classical information theory [73].

4.2.1 Objective function formulation

In this thesis, two criteria: selectivity and diversity have been considered in objective function formulation. In this formulation, there are M potential sensors to identify N *vapors.* We need to optimize a subset of size K out of M potential sensors ($K \leq M$) where, S_{ij} denotes the sensitivity of sensor j towards target vapor i. The objective function 1 has been formulated as follows:

Step 1: Normalization of matrix S_{ij} with respect to the total sum of elements in jth column:

$$\tilde{S}_{ij} = \frac{S_{ij}}{\sum_{i=1}^{N} S_{ij}} \tag{4.1}$$

where i and j refer to the index of analyte and sensor, respectively.

Step 2: Calculation of entropy, $H(\tilde{S}_j)$ for each selected sensor j:

$$H(\tilde{S}_j) = \frac{\log \frac{1}{\prod_{i=1}^{N} \tilde{S}_{ij}{}^{\tilde{S}_{ij}N}}}{K} \tag{4.2}$$

Step 3: Getting the maximum entropy (H_{max}) for sensor array of size K:

$$H_{max} = \log K \tag{4.3}$$

Step 4: Finally, the entropy-based objective function for a sensor array of size K is obtained as:

$$f_1 = \frac{1}{K} \sum_{j=1}^{K} \frac{H\,(\tilde{S}j)}{H_{max}} \tag{4.4}$$

Similarly, the objective function 2 has been formulated as follows:

Step 1: Normalization of matrix Sij with respect to the total sum of elements in ith row:

$$\tilde{S}_{ij} = \frac{S_{ij}}{\sum_{j} S_{ij}} \tag{4.5}$$

where, i and j refer to the index of analyte and sensor, respectively, as before.

Step 2: Calculation of entropy, $H(\tilde{S}_i)$ for each vapor class i:

$$H(\tilde{S}_i) = \frac{\log \frac{1}{\prod_{j=1}^{k} \tilde{S}_{ij}{}^{\tilde{S}_{ij}N}}}{K} \tag{4.6}$$

Step 3: Getting the maximum entropy (H_{max}) of the classification task with N vapor classes:

$$H_{max} = \log N \tag{4.7}$$

Step 4: Finally, the entropy-based objective function of the classification task with N vapor classes:

$$f_2 = \frac{1}{N} \sum_{i=1}^{N} \frac{H\,(\tilde{S}i)}{H_{max}}; \tag{4.8}$$

In this thesis, weighted sum approach has been used, taking w be the weight to solve the deterministic multi-objective problem. Therefore, our objective function becomes:

$$min\, f\ (f_1, f_2) = w \times \frac{1}{K} \sum_{j=1}^{K} \frac{H\ (\tilde{S}j)}{H_{max}} + (1\text{-}\ w) \times \frac{1}{N} \sum_{i=1}^{N} \frac{H\ (\tilde{S}i)}{H_{max}} \tag{4.9}$$

4.2.2 Constraints formulation

In this thesis, the constraints have been developed based on the limitation proposed by Shannon in his information theory [73]. Here, the number of constraints depend on the number of sensors used and the number of analytes being analyzed. For a number of i analytes and j number of sensors, there will be $(i + j)$ number of constraints. the constraints are formulated as follows:

$$g_j = \left|\frac{\log {S_j}^{-1}}{K} - \log K\right| < 0.1;\ \text{for } j = 1, 2, 3\ldots\ldots K \tag{4.10}$$

$$g_i = \left|\frac{\log {S_i}^{-1}}{N} - \log N\right| < 0.1;\ \text{for } i = 1, 2, 3\ldots\ldots N \tag{4.11}$$

4.3 DO Model Formulation:

Design variables and system parameters are considered fixed in the deterministic optimization formulation. The randomness in system parameters and uncertainty in obtaining precise design variables are not considered in this type of formulation. Using Eqs. (4.9) to (4.11), the deterministic optimization is formulated as:

$$min\, f = w \times \frac{1}{K} \sum_{j=1}^{K} \frac{H\ (\tilde{S}j)}{H_{max}} + (1\text{-}\ w) \times \frac{1}{N} \sum_{i=1}^{N} \frac{H\ (\tilde{S}i)}{H_{max}} \tag{4.12}$$

$$s.t.\quad \left|\frac{\log {S_j}^{-1}}{K} - \log K\right| < 0.1;\ \text{for } j = 1, 2, 3\ldots\ldots K \tag{4.13}$$

$$\left|\frac{\log {S_i}^{-1}}{N} - \log N\right| < 0.1;\ \text{for } i = 1, 2, 3\ldots\ldots N \tag{4.14}$$

$$\boldsymbol{lb} \leq \boldsymbol{S_{ij}} \leq \boldsymbol{ub} \tag{4.15}$$

Here, f is the objective function, $\boldsymbol{S_{ij}}$ is the vector of design variables, and $\boldsymbol{lb}$ and $\boldsymbol{ub}$ are the vectors of lower and upper bounds of design variables. In practice, the design variables and system parameters might be uncertain and solutions obtained from deterministic formulation can be sensitive due to the unaccounted uncertainty.

4.4 RDO Model Formulation:

In robust-based design optimization, the uncertainty present in the design variables is considered. In this thesis, the aleatory uncertainty present in the sensitivities of the sensors has been considered. The mean values of the sensitivities are obtained from the average of multiple observations and 0.1 is considered as the standard deviation as suggested by Xu et al. [3]. By applying first order Taylor series expansion method, the mean and variance of the objective function and constraints can be obtained [46], respectively, using Eqs. (4.12) and (4.13) as follows.

$$\mu_f = \mathrm{E}\,(f) = \frac{1}{K}\sum_{j=1}^{K}\frac{H\,(\tilde{S}j)}{H_{max}}; \tag{4.16}$$

$$\sigma_f{}^2 = \sum_{j=1}^{K}\left(\frac{\partial\,(f)}{\partial S_{ij}}\right)^2 Var(S_{ij}) \tag{4.17}$$

In this research, we have simultaneously minimized the mean and standard deviation of the first objective function which is based on selectivity criterion. The second objective function is converted to a constraint using the ε-constraint method using Eq. (3.22) [74]. Our robustness based design optimization model can be written as:

$$\min_{S_{ij}} f\,(\mu,\sigma) = w \times \frac{1}{K}\sum_{j=1}^{K}\frac{H\,(\tilde{S}j)}{H_{max}} + (1\text{-}w) \times \sqrt{\sum_{j=1}^{K}\left(\frac{\partial\,(f)}{\partial S_{ij}}\right)^2 Var(S_{ij})} \tag{4.18}$$

$$s.t.\ \boldsymbol{LB}+k\sigma\left(\frac{\log S_j{}^{-1}}{K} - \log K\right) \le E\left(\frac{\log S_j{}^{-1}}{K} - \log K\right) \le UB - k\sigma\left(\frac{\log S_j{}^{-1}}{K} - \log K\right) \tag{4.19}$$

$$\boldsymbol{LB}+k\sigma\left(\frac{\log S_i{}^{-1}}{N} - \log K\right) \le E\left(\frac{\log S_i{}^{-1}}{N} - \log K\right) \le UB - k\sigma\left(\frac{\log S_i{}^{-1}}{N} - \log K\right) \tag{4.20}$$

$$\boldsymbol{lb_{ij}} + k\sigma_{s_{ij}} \le \boldsymbol{E}\,(\boldsymbol{S_{ij}}) \le \boldsymbol{ub_{ij}} - k\sigma_{s_{ij}} \tag{4.21}$$

where, the mean, μ_f and standard deviation, σ_f values of the objective function is obtained from Eq 4.12 and Eq. 4.13, respectively. $\boldsymbol{E}\,(\boldsymbol{S_{ij}})$ is the mean values of the uncertain design variables; $i\times j$ is the number of random design variables; $w \geq 0$ is the weighting coefficients that represent the relative importance of the objectives; ***LB*** and ***UB*** are the vectors of lower and upper bounds of the inequality constraints; ***lb*** and ***ub*** are the vectors of lower and upper bounds of the design variables. k =1 is used here to adjust the robustness of the method against the level of conservatism of the solution. Considering the variations in the design variables, k reduces the feasible region and is related to the probability of constraint satisfaction.

4.5 RBDO Model Formulation:

In this thesis, the FORM based RBDO is used. In FORM approach, failure probability for any limit state function can be estimated using three steps. These are as follows—

Step 1: Transformation of the random variables $\boldsymbol{r}$ to the standard normal space $\boldsymbol{u}$, such that:

$$u_{ij} = \frac{r_{ij} - \mu_{ij}}{\sigma_{ij}} \text{ ; for all } i, j \tag{4.22}$$

Step 2: Calculation of the most probable point (MPP) of failure, $\boldsymbol{u*}$. This point is the solution to the constrained optimization problem—

$$\boldsymbol{u}^* = \arg\min (\| u \| \; g(u) = 0) \tag{4.23}$$

where, $\|\boldsymbol{u}\|$ is the Euclidean norm of vector $\boldsymbol{u}$; $g(\boldsymbol{u})$ represents a limit state function.

Step 3: Calculation of the reliability index β. For most practical problems, β is greater than zero, in which case β is also equal to $\|\boldsymbol{u}^*\|$ The probability of failure is approximated as: $P_f = \Phi(-\beta)$, where Φ is the standard Gaussian cumulative distribution function (CDF).

The main challenge in RBDO lies in handling the reliability constraints stated in Eq. (3.9). RBDO methods can be classified into 3 groups based on how reliability analysis is handled in the optimization process. These are— (i) nested double loop methods; (ii) single loop methods and (iii) decoupled methods. RBDO based on FORM is a nested double loop approach that requires an optimization loop, as well as a reliability loop, i.e., two optimizations, are coupled. Such coupling of the two optimization loops is computationally expensive. Several single loop methods and decoupled methods have been developed to reduce the computational burden of the nested approach. RBDO based on FORM can be classified into two approaches— Reliability Index Approach (RIA) and Performance Measure Approach (PMA). In this thesis, RBDO based on FORM is used that follows RIA. Therefore, the RBDO formulation stated by Eqs. (3.9) to (3.12), can be reformulated as follows:

$$min\, f\ (f_1, f_2) = w \times \frac{1}{K}\sum_{j=1}^{K}\frac{H\ (\tilde{S}j)}{H_{max}} + (1\text{-}\ w) \times \frac{1}{N}\sum_{i=1}^{N}\frac{H\ (\tilde{S}i)}{H_{max}} \tag{4.24}$$

$$s.t.\ \ P_{fq} = \phi(\text{-}\beta_q) \leq \mathrm{P}_{fq}{}^{allowable}$$

$$\text{where, } \beta_q = \|\boldsymbol{u_q}^*\|;\text{ and }\ \boldsymbol{u_q}^* = \arg\min\ (\|\boldsymbol{u}\|\ \mathrm{g}_q\ (u) = 0)\text{ for all } q \tag{4.25}$$

$$\boldsymbol{lb} \leq \mu_{ij} \leq \boldsymbol{ub} \tag{4.26}$$

where, f is the objective function; μ_{ij} is the mean values of the random design variables (S_{ij}); $\boldsymbol{u}$ is transformed standard normal space vector corresponding to the vector of all the random variables $\boldsymbol{r}$; $g_q(\boldsymbol{u})$ is the qth limit state function expressed in standard normal space; β_q is the reliability index for qth limit state function; Φ is the standard Gaussian cumulative distribution function (CDF); P_{fq} and $P_{fq}{}^{allowable}$ represent the failure probability and allowable (maximum) failure probability for qth limit state function, respectively; $\boldsymbol{\mu_q}^*$ is the most probable point (MPP) of failure for qth limit state function; the vectors $\boldsymbol{lb}$ and $\boldsymbol{ub}$ represent lower and upper bounds of all the design variables, respectively. The probabilistic formulation of Eq. (3.10) is substituted by Eq. (4.25). The RBDO formulation stated by Eqs. (4.24) to (4.26), is used to solve the sensor array optimization problem. RBDO used in this thesis is formulated to deal with aleatory uncertainty only; epistemic uncertainty is out of the scope of this thesis.

4.6 Evaluation of the Optimization Results

4.6.1 General Resolution Factor (GRF)

Several criteria have been proposed for evaluating the performance of the selected sensors in the literature such as the classification rate [75], the distance measure [1, 3], and the comparison between classification rate and distance measure [76]. Considering that the classification rate does not only depend on the quality of input features but also on the parameter setting and the environment configuration, the evaluation result is produced by employing a classifier that does not guarantee the quality of the input features. Therefore, a high correct classification rate can mean overfitting. According to the above explanation, it will be necessary to verify the quality of the input features before they are consumed by another process. This work uses General Resolution Factor (GRF) to measure the input quality of the selected features. GRF can be expressed by the following equation [1]:

$$GRF = \sqrt{\sum_{i=1}^{m}\frac{(\mu_{i1} - \mu_{i2})^2}{\sigma_{i1}^2 + \sigma_{i2}^2}} \tag{4.27}$$

where, i is the index of feature; m is an integer and $d \geq 1$, it equals to the total number of input feature(s); μ_{i1} and μ_{i2} are group mean values, and σ_{i1} and σ_{i2} are standard deviations of the two classes with respect to feature i. Assuming Gaussian distribution, the larger ratio between centroid distance $\sqrt{\mu_{i1} - \mu_{i2}}$ and $\sqrt{\sigma_{i1}^2 + \sigma_{i2}^2}$ is related with a larger probability of correct classification rate [77]. Therefore, higher GRF value indicates better performance for a sensor group.

4.6.2 Principal component analysis

In this thesis, the input quality of each sensor group is demonstrated by Principal Component Analysis (PCA) plot. After the projection of original data, the resulting space can best explain the variation of original data in a sum-squared error sense [78]. Principal components (PCs) are new variables generated through linear transformation; and the axes corresponding to PCs are orthogonal to each other. Principal components are numbered according to the amount of variance in original data they account for. The first PC explains the maximum variance, the second PC accounts for the maximum variance that is not covered by the first PC, and so on [71]. The eigenvalues corresponding to those principal components are also in decreasing order. Usually, a few first PCs with the largest eigenvalues compose of the subspace dominating the original signal, while the remaining PCs contain noise [78]. PCA can reduce dimension, and eliminate noise through keeping only several first principal components. Two-dimensional and three-dimensional plots of dominating principal components are common tools for visual inspection of classification. In this research, two-dimensional PCA plot is used to evaluate the performance of the selected sensors graphically.

The performance of the optimized set of sensor arrays are evaluated by the General Resolution Factor (GRF). Generally, the higher value of GRF indicates the better performance of the sensor array. Moreover, the classification performance of different sensor groups are evaluated with the aid of the Principal Component Analysis (PCA). The input quality of the selected sensors can be visually evaluated by their separation of the Principal Components (PCs). The following chapter provides a numerical illustration of the proposed method with results and discussions.

CHAPTER 5
NUMERICAL ILLUSTRATIONS

5.1 Sensor System Setup

A sensor array system includes several sensor elements each coated with certain sensing material which has particular sensitivity to analytes, such as vapors, smokes and different volatile organic compounds (VOCs). In this thesis, a Message Queue (MQ) based sensor system prototype is built for demonstrating the proposed model. The MQ sensors are low-cost and reliable sensors that can detect various kinds of gases, smokes and particulates. Our sensor system is built based on the Arduino platform. The sensor system is designed to test the presence of different gases. The sensor system used nine different MQ-based sensors. In Table 5.1, the types of MQ sensors used in this study are shown with their detecting capability.

Table 5.1: Name of the MQ Sensors with their Selectivity

Serial No	Sensor	Selectivity
1	MQ2	LPG, Propane, Methane, Alcohol, Hydrogen, Smoke
2	MQ3	Alcohol, Benzene, Methane (CH_4), Hexane, LPG, CO_2
3	MQ4	Methane (CH_4), CO and Natural gas
4	MQ5	LPG, Natural gas
5	MQ6	LPG, iso-butane, Propane
6	MQ7	Carbon Monoxide (CO)
7	MQ8	Hydrogen (H_2)
8	MQ9	Methane (CH_4), Propane and CO_2
9	MQ135	Ammonia (NH_3), NO_x, Alcohol, Benzene, Smoke and CO_2

The sensor system consists of an Arduino UNO R3, nine different sensors as shown in Table 5.1, breadboard, red and green LEDs, resistors, buzzer and different required wires. The open-source Arduino Software (IDE)was used to compile and upload the code to the UNO R3 board. The sensor system prototype is shown in Figure 5.1.

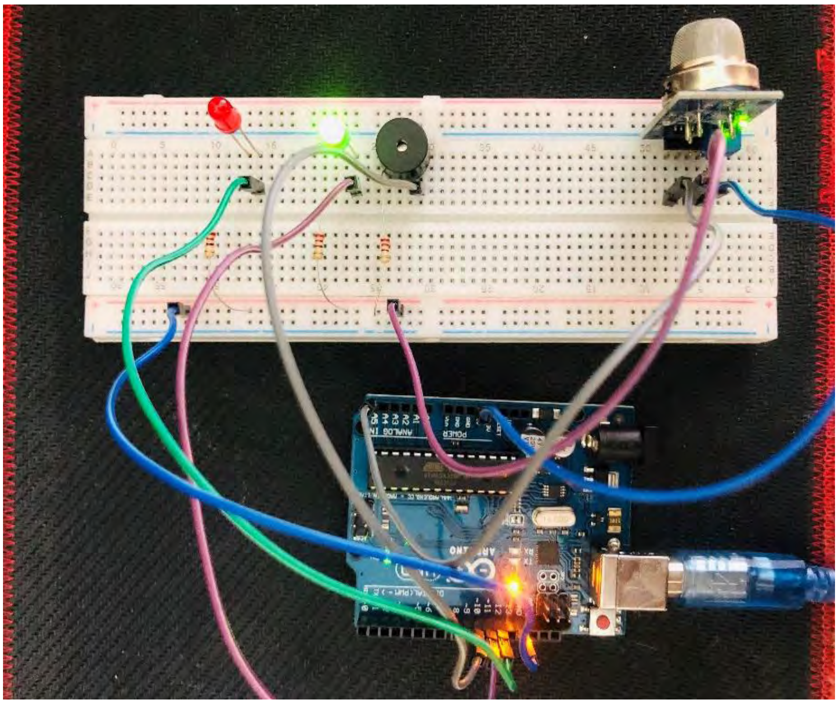

Figure 5.1: Sensor System Setup

5.2 Calculation of Sensitivity

The steady-state of sensor response is calculated from the relative differential resistance change. Usually, the relationship between sensor response and concentration is quasi-linear in some range of concentration. Therefore, the response of sensor i to the analyte j can be estimated as:

$$R_{ij} \cong S_{ij}C_j \tag{5.1}$$

where, S_{ij} is the sensitivity of the ith sensor to the jth analyte, C_j is the concentration in Parts Per Million (PPM). Here, sensitivity summarizes the essential attribute of senor response. In this study, sensitivity was used as the basis for sensor selection. In the sensor system, the vapors tested were: Carbon monoxide (CO), Liquefied Petroleum Gas (LPG), and Smoke. Generally, gas sensor is an analog sensor so the output data are the result of analog to digital conversion (ADC). The ADC value must be converted to obtain the resistance value of the sensor from Eq. (5.2) and Eq. (5.3).

$$Rs_i = \frac{V_c - V_{RLi}}{V_{RLi} \times RL_i} \tag{5.2}$$

$$V_{RLi} = \frac{ADC_i - V_c}{C_{byte}} \tag{5.3}$$

where, R_s is sensor resistance at various concentrations; i is the number of sensors; V_c is the circuit voltage (5V±0.1); V_{RL} is the voltage sensor in the sample space; RL is sensor load resistance; ADC is Analog to Digital Conversion values from each sensor; C_{byte} : Byte of the used board (1024 byte). In this thesis, the value of sensor response and the concentration of the vapor in ppm are obtained from the readings from the Arduino Uno which is shown in the Arduino Software's output interface to calculate the sensitivities of each sensor to each analyte.

5.3 Cluster Analysis

The data sets are primarily analyzed by cluster analysis. Euclidean distance is used as similarity measurement which gives similar grouping results. In hierarchical clustering, the dendrogram is used to allocate objects to clusters or groups based on the similarity measurement [78]. User can choose the number of clusters or groups according to his preference based on the similarity measure. In this thesis, we have chosen five clusters or groups. In Figure. 5.2, the dendrogram indicates that the 9 sensors could be roughly divided into different groups by cutting at a different level of similarity. Figure 5.2 shows that the grouping can be done as 3 sensors group, 4 sensors group, 5 sensors group and 7 sensors group by cutting the dendrogram at 97.5, 98.5, 99 and 99.5 percent similarity, respectively. We have also taken all 9 sensors in a group for comparison with other groups.

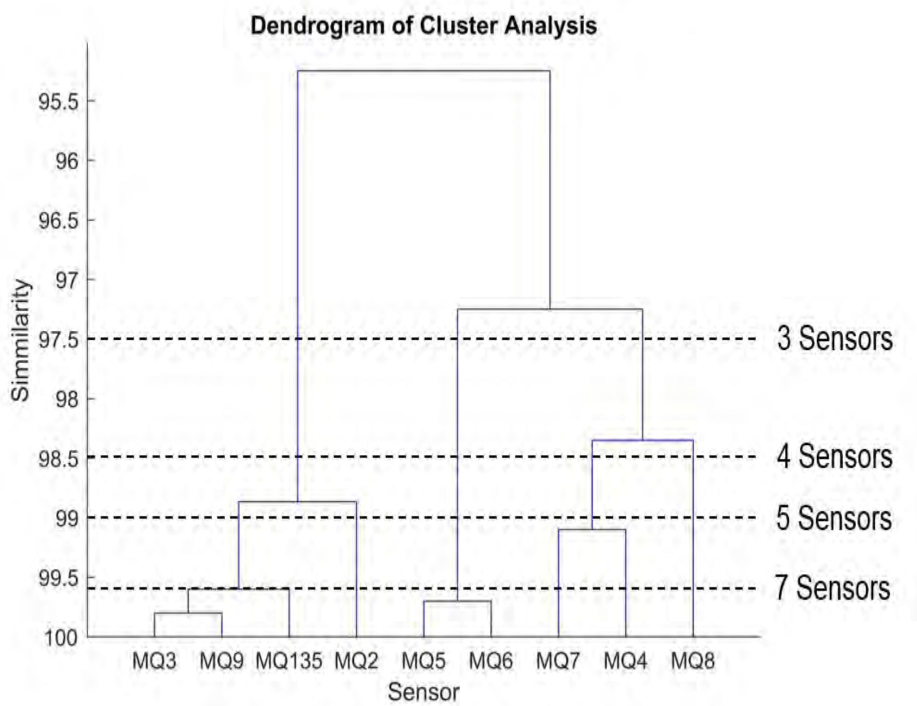

Figure 5.2: Dendrogram for Cluster Analysis

The members of each groups are selected by the modified Fast Correlation-Based Filter (FCBF) proposed by Wijaya et al. [79]. The FCBF algorithm can find the best combination of sensor array with minimizing feature redundancy. The sensor groups with their selected member are shown in Table 5.2.

Table 5.2: Sensor Groups after Cluster Analysis

Sensor Array	Sensor Members
3 Sensors	MQ4, MQ5, and MQ135
4 Sensors	MQ4, MQ5, MQ8 and MQ135
5 Sensors	MQ2, MQ4, MQ5, MQ8, and MQ135
7 Sensors	MQ2, MQ3, MQ4, MQ5, MQ7, MQ8, and MQ135
All 9 Sensors	MQ2, MQ3, MQ4, MQ5, MQ6, MQ7, MQ8, MQ9, and MQ135

5.4 Results and Discussion

After the cluster analysis, the deterministic, robust, and reliability-based multi-objective optimizations were realized in the MATLAB environment and run on a computer with Intel Core i5 processor, and 8 GB RAM. The results obtained from each of the formulations are discussed in the following sections.

5.4.1 Results from deterministic formulation

First, the deterministic problem formulated in Eqs. (4.12) to (4.15) has been solved. The weighted sum approach has been used to solve the multi-objective problem. The weight is taken from 0 to 1 with an increment of 0.1. The *fmincon* solver was used to solve the problem. The Pareto fronts obtained for different sensor groups of Table 5.2 using the deterministic formulation are shown in Figure 5.3.

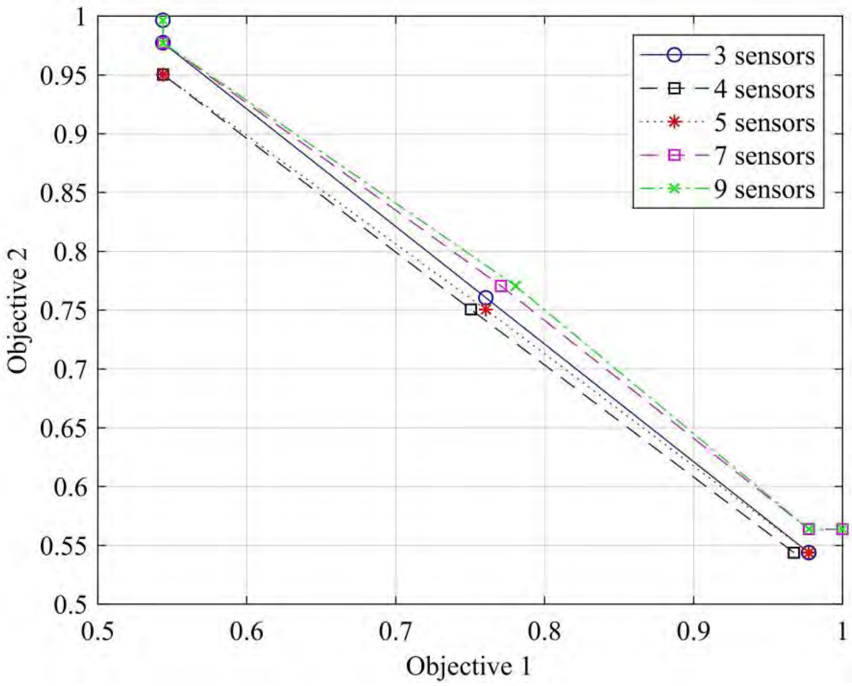

Figure 5.3: Pareto Optimal Solutions for the deterministic problem for different sensors group

As the weighting coefficient (w) increases, the value of objective function 1 decreases and objective function 2 increases. As the problem is formulated as minimization problem, lower values for both objective function 1 and objective function 2 are desirable. From Figure 5.3, it can be observed that in deterministic formulation, for 4 sensors group the value for both objective functions is lower compared to other groups. For 3 sensors group and 5 sensors group, the value of objective functions is less than the value for 7 sensors group and 9 sensors group. It can be observed that, with the increase in number of sensors from 4 sensor group to 9 sensor group, the performance of the sensor groups is degrading. Therefore, it can be concluded that the 4 sensors group can be chosen from the DO optimization. In addition, the performance of 3 sensors group and 5 sensor group is better than 7 sensors group and 9 sensors group. Specially, the performance significantly reduced in case of the 9 sensors group because the values of both objective functions are found to be higher than the other groups. In 9 sensors group, the greater number of sensors is responsible for increased complexity leading to higher noise. The generation of more noise results in higher entropy which is responsible for their performance degradation.

5.4.2 Results from robustness-based formulation

The RDO formulation presented in Eq. (4.16) to Eq. (4.21) of Chapter 4 is solved with multiple combinations of the weighting coefficients to obtain Pareto front for the multi-objective optimization problem using MATLAB. We have used the *fmincon* solver for this purpose. The Pareto optimal solutions are obtained by varying w from 0 to 1 with an increment of 0.1. The Pareto fronts obtained for different sensor groups from the robustness-based formulation are shown in Figure 5.4.

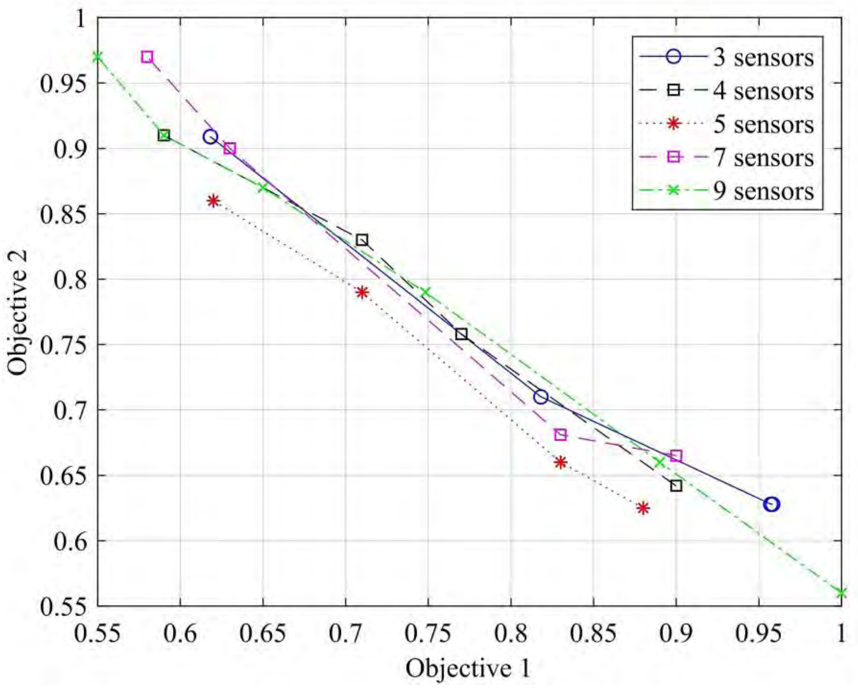

Figure 5.4: Pareto Optimal Solutions for the robustness-based problem for different sensors group

As the weighting coefficient (w) increases, the value of value of objective function 1 decreases and objective function 2 increases. As the problem is formulated as minimization problem, lower values for both mean and standard deviation of the objective function are preferable. From Figure 5.4, it can be observed that the performance of 5 sensors group is best for having lowest values for both objective functions. The performance of 4 sensors group, 3 sensors group and 5 sensors group can be considered as second, third, and fourth respectively. It can be observed that the performance degrades in either increasing or decreasing the number of sensors from 5 sensors groups. The performance of all 9 sensors together is poor because it fails to provide lower value for both the objective functions.

5.4.3 Results from reliability-based formulation

The reliability-based design optimization (RBDO) formulation presented in Eq. (4.22) to Eq. (4.26) is solved using MATLAB. This problem is also solved using the *fmincon* solver of MATLAB. We have used the FORM method following RIA to solve the reliability-based method. The Pareto optimal solutions are obtained by varying w from 0 to 1 with an increment of 0.1. The optimization loop includes reliability constraints that require prior evaluation of the reliability index (β) through additional optimization processes for each limit state function or inequality constraint. The optimization processes for the evaluation of β are executed in terms of standard normal vectors $\boldsymbol{u}$ corresponding to the random vector $\boldsymbol{r}$. No lower or upper bounds are imposed on $\boldsymbol{u}$. The nested optimizations require higher computational effort compared to RDO approaches. The Pareto front obtained in RBDO formulation is depicted in Figure 5.5.

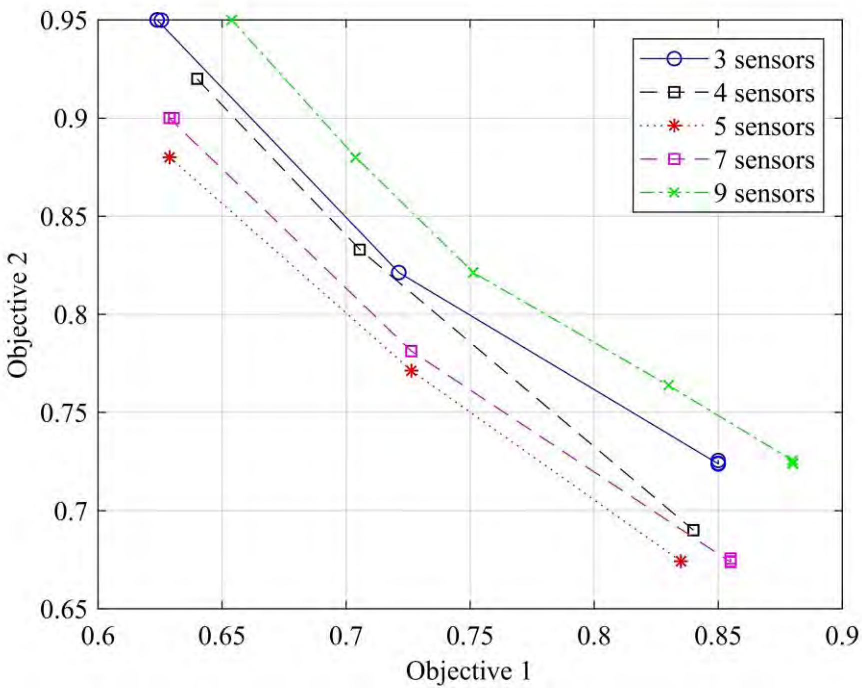

Figure 5.5: Pareto Optimal Solutions for the reliability-based problem for different sensors group

As the weighting coefficient (w) increases, the value of objective function 1 decreases and objective function 2 increases. As the problem is formulated as minimization problem, lower values for both objective functions are desirable. From the figure 5.5, it can be observed that the performance of 5 sensors group is optimal for providing lower values for both the objective functions. The performance of 4 sensors group and 7 sensors group can be considered in next place after 5 sensors group. And, the performance of the 3 sensors group and 9 sensors group is poor compared to the other groups for having higher values for both the objective functions. Therefore, it can be said that the performance degrades in either increasing or decreasing the number of sensors from the optimal 5 sensors group. The absence of required number of sensors in 3 sensor group and presence of redundant sensors in 9 sensors group might be responsible for their performance degradation, respectively.

5.4.4 Discussions of results

Pareto optimal solutions to the DO, RDO, and RBDO formulations provide multiple alternative sets of optimal values for the sensitivities to select optimal set of sensors for sensor array. The computational time of the different methods used in this thesis are compared in Table 5.3, for different sensors groups with the weighting coefficient, w = 0.5.

Table 5.3: Computational time in seconds for deterministic, robustness, and reliability-based formulations

Sensors Group	Deterministic Optimization	Robustness-based optimization	Reliability-based optimization
3 Sensors Group	1.22 sec	2.18 sec	118.11 sec
4 Sensors Group	1.31 sec	2.25 sec	149.53 sec
5 Sensors Group	1.45 sec	2.34 sec	181.19 sec
7 Sensors Group	1.81 sec	2.55 sec	221.20 sec
9 Sensors Group	1.97 sec	3.04 sec	245.77 sec

The Pareto fronts obtained in different methods are illustrated in Figure 5.6 to compare the values of the two objective functions corresponding to different Pareto optimal solutions for 3 sensors group, 4 sensors group, 5 sensors group, 7 sensors group and 9 sensors group, respectively.

The Pareto optimal solutions offer a set of alternative combinations to be chosen according to the requirements of the user. The user can choose one of the Pareto optimal solutions and select the optimal sensor set for designing a sensor system. It can be observed from Figure 5.6 (a) that for 3 sensors group, the objective function values provided by the RDO are preferable compared to those of DO and RBDO. In addition, from Table 5.3, it can be observed that the computational time involved in RDO approaches is lower compared to

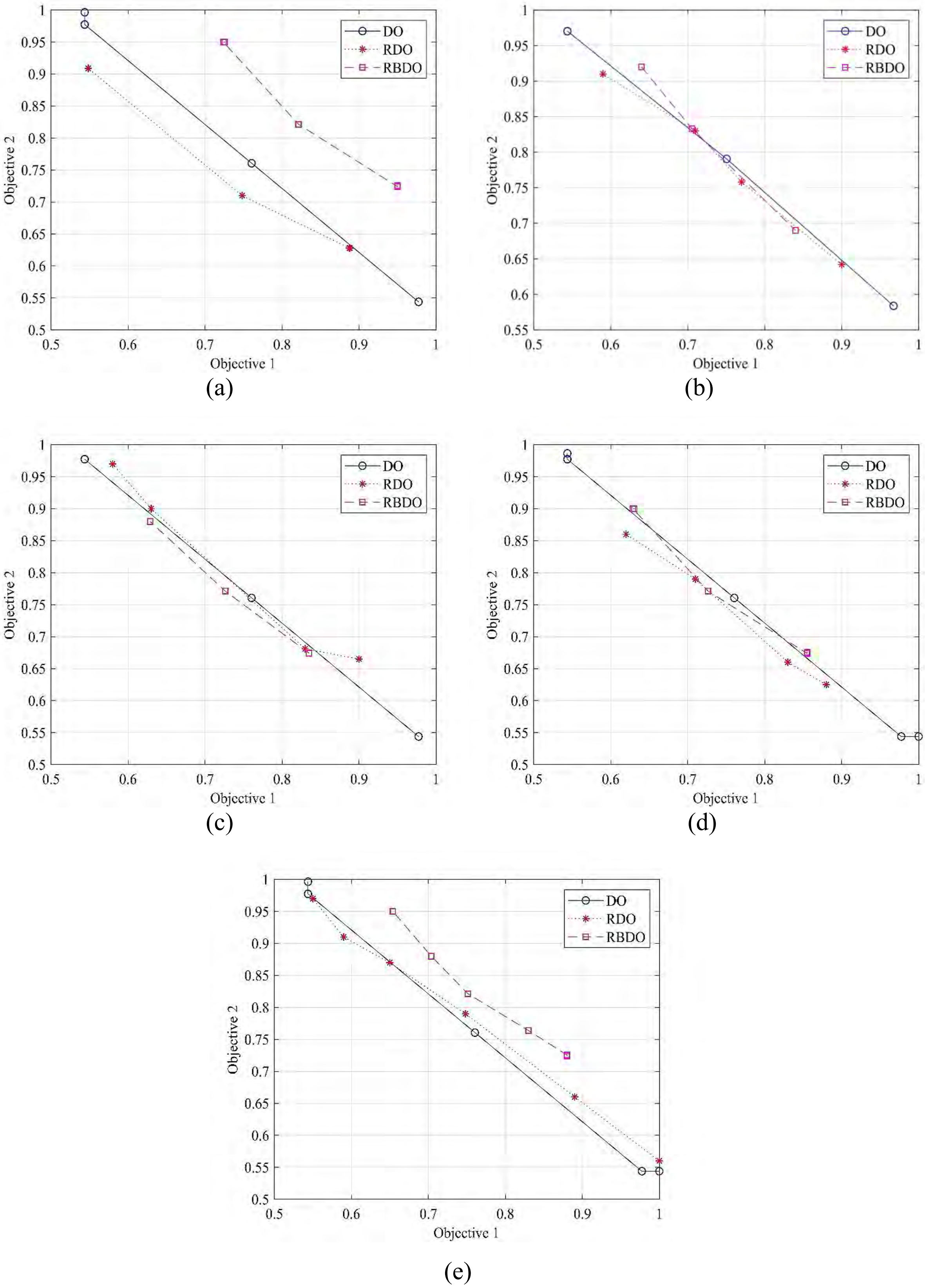

Figure 5.6: Comparison among the Pareto Optimal Solutions obtained in DO, RDO and RBDO formulations for: (a) 3 sensors group (b) 4 sensors group (c) 5 sensors group (d) 7 sensors group and (e) 9 sensors group

RBDO approaches but higher than DO. From figure 5.6 (b), it can be observed that the performance of RDO and RBDO is almost similar but better than DO. In this case, DO is taking less time than RDO and RBDO. Again, It can be observed from Figure 5.6 (c) that for 5 sensors group the objective function values provided by the RBDO are preferable compared to those of DO and RDO. In contrast, the computational time involved in RBDO approaches is significantly higher compared to DO and RDO approaches. From Figure 5.6 (d), for 7 sensors group, the objective function values provided by the RDO are preferable compared to those of DO and RBDO. In this case, the computational time involved in RDO approaches is lower compared to RBDO approaches but higher than DO. From Figure 5.6 (e), it can be observed that DO performed better in minimizing the objective functions in case of 9 sensors group compared to RDO and RBDO. From the Table 5.3, it can be observed that the computational time involved in DO approaches for 9 sensors group is lower compared to RDO and RBDO approaches.

As the number of sensors increases, the computational complexity increases. Therefore, the computational time also increases for every formulation with higher number of sensors. By comparing the results for different sensors group, DO is found to be taking less computational time in all cases. RBDO is found to be taking more time in all cases due to presence of extra reliability loop as well as optimization loop. RDO lies in between DO and RBDO for all sensors groups. However, the performance of DO, RDO and RBDO in achieving lower values for objective functions varies for different sensors groups. It is recommended to use RDO for 3 sensors group, 4 sensor group, and 7 sensors group, RBDO for 5 sensors group and DO for 9 sensors groups.

5.5 Evaluation of the Optimization Results

As mentioned earlier, the performance of the selected sensors based on the optimization results are evaluated using General Resolution Factor (GRF). Quantitatively, the feature subset quality of each sensor group is compared based on GRF. First, the original space for each sensor combination is transformed into a new space by the help of PCA. GRF values are calculated by the first two of the corresponding principal components (PCs) from each sensor group. Table 5.4 summarized the GRF values obtained from each sensor groups.

Table 5.4: GRF values of Each Group

Sensor Group	GRF Value
3 Sensors	0.23901
4 Sensors	0.25417
5 Sensors	0.26781
7 Sensors	0.24925
9 Sensors	0.22766

From Table 5.4, it can be observed that the 5 sensors group has the highest GRF value than others. As the higher GRF value is related with a larger probability of correct classification rate [79]. Therefore, the performance of the 5 sensors can be considered as better than other groups. The GRF values for 4 sensors group and 7 sensors group are close to the 5 sensors group. The performance of the 4 sensors group and 7 sensors group stand next to the 5 sensors group. The GRF value is the lower for 3 sensors group compared to 4, 5 and 7 sensors group, thus, can be considered having poor performance than the 4, 5, and 7 sensors groups. And, the GRF value is lowest for using all 9 sensors together. Therefore, the performance of the 9 sensors group can be considered as poorest for having lowest GRF value among the five groups.

The input qualities for pattern recognition provided by each sensor groups can be visually evaluated by the Principal Component Analysis [3]. In this thesis, 2-D PCA plots have been used the evaluate the performance of the sensor groups. The PCA plots for the different sensor groups are shown in Figure 5.7.

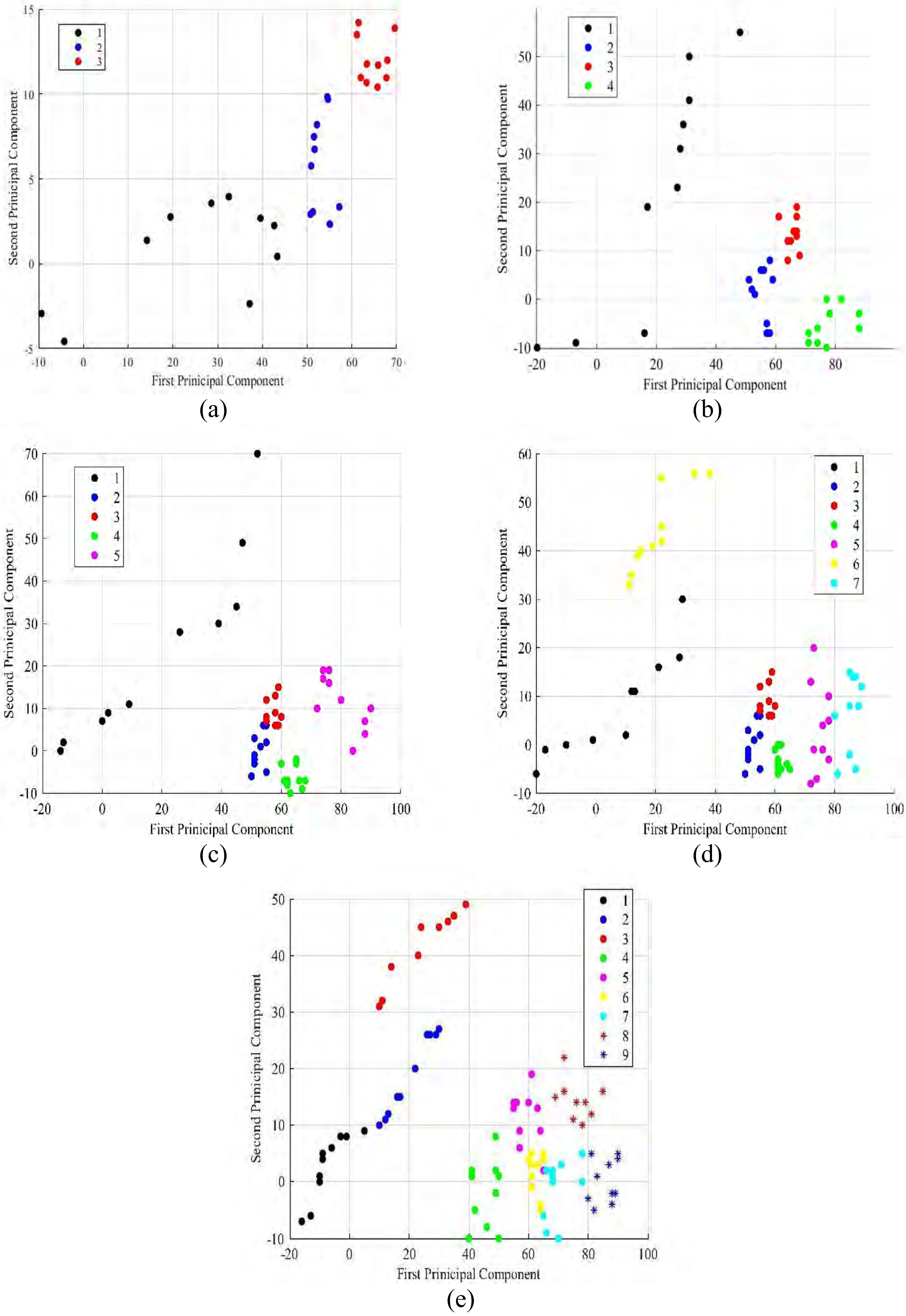

Figure. 5.7: PCA plot for sensor groups: (a) 3 sensors group (b) 4 sensors group (c) 5 sensors group (d) 7 sensors and (e) 9 sensors

PCA plots in Figure 5.7 indicate that all sensor groups are able to differentiate different vapor classes. The observations for 3 sensors group, 4 sensors group, and 5 sensors group show that these groups are able to separate the vapor classes without overlapping. The performance of the 5 sensors group is better in separating the vapor classes. Though, the 7 sensors group shows some overlapping, but is able to differentiate vapor classes satisfactorily. However, the overlapping is present to a large extent for 9 sensors group. The overlapping is caused by the presence of redundant sensors in 9 sensors group which is also responsible for their performance degradation.

The outcome of the proposed method is to select suitable subset of sensors from a set of sensors. In this thesis, total 9 sensors have been used to demonstrate the proposed methodology. After cluster analysis, five groups are considered consisting 3 sensors, 4 sensors, 5 sensors, 7 sensors, and 9 sensors, respectively. Our goal is to select the optimal set among these five groups. The deterministic optimization results shows that the performance of 4 sensors group is better. However, the results obtained from the deterministic optimization may fail due to the uncertainties present in the input variables. Optimization under uncertainty such as RDO and RBDO consider these uncertainties present in the input variables and provide more robust and reliable solutions. In our research, both robustness and reliability-based optimization results show that 5 sensors group is the optimal subset of sensors from the set of 9 sensors. The results obtained from RDO and RBDO considering the uncertainties in input variables is more reliable in real life scenario. Moreover, the GRF value is also higher for the 5 sensors group, and thus, supports the selection of 5 sensors group based on stochastic optimization results provided by RDO and RBDO. The performance of the 5 sensors group is visually evaluated by its classification performance using PCA. In the PCA plots, the 5 sensors group showed better separation with lowest overlapping than other sensor groups. Moreover, our optimization results from DO, RDO and RBDO shows that the performance of using all 9 sensors is the poorest among all other groups. The 9 sensors group has also the lowest GRF and highest overlapping in PCA plot, therefore indicating poorest performance among all groups. Therefore, using all 9 sensors together is not recommended at all. At least 2 redundant sensors need to be removed from the 9 sensors before using. However, using the 5 sensors group will be the optimal choice with a less chance of failure under uncertainties for getting improved input qualities for the pattern recognition part.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

Multi-objective sensor array optimization under uncertainty can significantly improve the input quality for the pattern recognition of the sensor systems. In the sensor system, successful detection is essential because it might be directly responsible for losing lives and loss of financial assets. By considering the aleatory uncertainty in the input variable, the proposed method can provide more robust and reliable solution in practical scenario. In this research, robustness and reliability-based multi-objective optimization methods have been proposed for sensor array optimization. The proposed methods are able to successfully determine the Pareto sets of potential optimal solutions; which could provide improved input quality for the pattern recognition. More importantly, selectivity and diversity were simultaneously used as two criteria in optimization to find optimal sensor arrays. Moreover, the proposed multi-objective optimization method under uncertainty was demonstrated using a designed sensor system prototype. The optimization results is evaluated quantitatively through the statistical measure of resolving power: general resolution factor (GRF) and visually with the aid of principal component analysis. The GRF values showed that most of the selected sensor arrays performed better than the combinations employing all potential sensors. The comparison of corresponding PCA plots also supported that argument. In general, this study provides the following contributions in the field of research:

i. Incorporation of uncertainty in the input variables to obtain more robust and reliable solutions for the sensor array optimization.
ii. Development of robust and reliability-based multi-objective sensor array optimization formulation considering both selectivity (sensor's response to the target analyte) and diversity (sensor's response to the rest of the analytes) criteria.
iii. Selection of the best combination of sensor array from a set of sensors based on the optimization results obtained from DO, RDO and RBDO.
iv. Removing redundant sensors from the sensor system which ensures that the industry can use a smaller number of sensors without sacrificing quality and become cost efficient.

Our proposed model of multi-objective optimization under uncertainty is capable of providing high quality input for pattern recognition with fewer sensors. The demonstration of the proposed methods using the sensor system prototype will facilitate the industries to understand and implement the proposed model easily in the practical arena. The application of this study is expected to significantly improve the sensor system design in the industrial sector.

6.2 Future Work

Several aspects of this research can be extended for the future development of sensor array optimization formulation. The determination of optimized location for placing these sensors would be a good addition. The sensor placement optimization will ensure the optimal locations for the sensors for better detection. Moreover, this thesis can be extended to consider the epistemic uncertainty. Therefore, future studies involving epistemic uncertainties and sensor placement optimization might improve the sensor system design.

REFERENCES

[1] Xu, Z., Shi, X., and Lu, S., "Integrated sensor array optimization with statistical evaluation," *Sensors Actuators, B: Chem.*, vol. 149, pp. 239–244, 2010.

[2] Zhang, W., Wang, L., Ye, L., Li, L., and Hu, M., "Gas sensor array dynamic measurement uncertainty evaluation and optimization algorithm," *IEEE Access*, vol. 7, pp. 35779-35794, 2019.

[3] Xu, Z., and Lu, S., "Multi-objective optimization of sensor array using genetic algorithm," *Sensors Actuators, B: Chem.*, vol. 160, pp. 278–286, 2011.

[4] Kira, K., and Rendell, L., "The feature selection problem: Traditional methods and a new algorithm," in *Proceedings of Ninth National Conference on Artificial Intelligence*, vol. 2, pp. 129–134, 1992.

[5] Jensen, R., and Shen, Q., "Computational intelligence and feature selection : rough and fuzzy approaches", chap. 1, pp. 5-7, John Wiley & Sons, Inc., Hoboken, New Jersey, 2008.

[6] Wang, L., Jia, P., Huang, T., and Duan, S., "A Novel Optimization Technique to Improve Gas Recognition by Electronic Noses Based on the Enhanced Krill Herd Algorithm.", *Sensors: MDPI*, vol. 16, pp. 1275-1290, 2016.

[7] Gualdrón, O., Llobet, E., Brezmes, J., Vilanova, X., and Correig, X., "Coupling fast variable selection methods to neural network-based classifiers: Application to multisensor systems," *Sensors Actuators B Chem.*, vol. 114, pp. 522–529, 2006.

[8] Gardner, J. W., Boilot, P., and Hines, E. L., "Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach," *Sensors Actuators B Chem.*, vol. 106, pp. 114–121, 2005.

[9] Bertsimas, D., Brown, D. B., and Caramanis, C., "Theory and Applications of Robust Optimization," *SIAM Rev.*, vol. 53, pp. 464–501, 2011.

[10] Granig, W., Faller, L. F., and Zangl, H., "Sensor system optimization to meet reliability targets," *Microelectronics Reliability*, vol. 87, pp. 113-124, 2018.

[11] Shi, S. G., Li, Y., Zhu, Z., and Shi, J., "Real-valued robust DOA estimation method for uniform circular acoustic vector sensor arrays based on worst-case performance optimization," *Applied. Acoustics.*, vol. 148, pp. 495–502, 2019.

[12] Feng, L., Musto, C.J., Kemling, J.W., Lim, S., Zhong, W., and Suslick, K., "Colorimetric sensor array for determination and identification of toxic industrial chemicals". *Analytical chemistry*, 82(22):9433–9440, 2010.

[13] Lubna, F., and Lu, S., “Colorimetric sensor array optimization using cluster analysis”. *In IIE Annual Conference. Proceedings*, pages 1835–1839. Institute of Industrial and Systems Engineers (IISE), 2017.

[14] Lin, H., and Suslick, K., A colorimetric sensor array for detection of triacetone triperoxide vapor. Journal of the American Chemical Society, 132(44):15519–15521, 2010.

[15] Feng, L., Musto, C.J., Kemling, J.W., Lim, S., Zhong, W., and Suslick, K., “A colorimetric sensor array for identification of toxic gases below permissible exposure limits”, *Chemical Communications*, 46(12):2037–2039, 2010.

[16] Puligundla, P., Jung, J., Ko, S., “Carbon dioxide sensors for intelligent food packaging applications”, *Food Control*. Vol. 25, pp.328–333, 2012.

[17] Falasconi, M., Concina, I., Gobbi, E., Sberveglieri, V., Pulvirenti, A., Sberveglieri, G., “Electronic Nose for Microbiological Quality Control”, *Food Products*, Vol. 2012, 2012

[18] McMeekin, T.A., Patterson, J.T. , “Characterization of hydrogen sulfide-producing bacteria isolated from meat and poultry plants”, *Applied Microbiology*. Vol. 29, pp.165–169,1975.

[19] Hasan, N., Ejaz, N., Ejaz, W., Kim, H., “Meat and fish freshness inspection system based on odor sensing”, *Sensors (Switzerland)*. Vol. 12, pp.15542–15557, 2012.

[20] Jia, P.F., Tian, F.C., Fan, S., He, Q, Feng, J.W., Yang, S., “A novel sensor array and classifier optimization method of electronic nose based on enhanced quantum-behaved particle swarm optimization”, *Sensor Review*. Vol. 34, pp.304–311, 2014.

[21] Kumar, A., Tudu, B., Roy, N., Bhattacharyya, R., “Optimization of Sensor Array” *Electronic Nose* :, Vol. 11, pp.3001–3008, 2011.

[22] Szecówka, P.M., Szczurek, A., Licznerski, B.W., “On reliability of neural network sensitivity analysis applied for sensor array optimization”, *Sensors and Actuators B: Chemical*. Vol. 157, pp.298–303, 2011.

[23] Saha, P., Ghorai, S., Tudu, B., Bandyopadhyay, R., Bhattacharyya, N., “Optimization of sensor array in electronic nose by combinational feature selection method”, *2012 Sixth International Conference on Sensing Technology (ICST)*. pp.341–346, 2012.

[24] Guyon, I., Gunn, S., Nikravesh, M., and Zadeh,L.A., “Feature extraction: foundations” *Sensor Applications*,36(10) 2006.

[25] N. Sánchez-Maroño, A. Alonso-Betanzos, M. Tombilla-Snaromán, Filter methods

for feature selection–a comparative study, Intelligent Data Engineering and Automated Learning - IDEAL 2007. pp.178–187, 2007. doi:10.1007/978-3-540-77226-2.

[26] Vergara, J., Este, P., "A review of feature selection methods based on mutual information", *Neural Computing and Applications*. 2013.

[27] Pradip, S., Ghorai, S., Tudu, B., Bandyopadhyay, R., and Bhattacharyya, N., "Optimization of sensor array in electronic nose by combinational feature selection method", *Sensing Technology: Current Status and Future Trends II*, pages 189–205. Springer, 2014.

[28] Sathishkumar, E.N., Thangavel, K., and Daniel, P., "Effective clustering algorithm for gas sensor array drift dataset." *Int. J. Comput. Intelli. Inform.*, 3: 144-150. 2018.

[29] Parkinson. A, Sorensen. C, and Pourhassan, N, "A general approach for robust optimal design" *Trans ASME*, 115:74–80. 2018.

[30] Du X, Chen, W, "Towards a better understanding of modeling feasibility robustness in engineering", *ASME J Meach Des*, 122(4):385–394. 2000.

[31] Doltsinis I, Kang Z Robust design of structures using optimization methods. Comput Methods Appl Mech Eng 193:2221– 2237, 2004.

[32] Huang B, Du X , Analytical robustness assessment for robust design. Struct Multidisc Optim 34:123–137, 2007.

[33] Chiralaksanakul A, Mahadevan, S., "First-order approximation methods in reliability-based design optimization". *J Mech Des*, 127(5):851–857, 2005.

[34] Ramu P, Qu X, Youn BD, Haftka RT, Choi KK, "Inverse reliability measures and reliability-based design optimization", *Int J Reliab. Saf.,* 1(1/2):187–205, 2006.

[35] Agarwal H, Renaud JE, Preston EL, Padmanabhan D., "Uncertainty quantification using evidence theory in multidisciplinary design optimization". *Reliability Engineering and System Safety*; 85(1–3):281–94, 2004.

[36] Steven D., "Robustness in Complex Systems". *Proceedings of the 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII),* 2001.

[37] Taguchi, G., "Taguchi on robust technology development: bringing quality engineering upstream", *ASME*, New York, 1993.

[38] Wei, D.L., Cui, Z.S., Chen, J, "Robust optimization based on a polynomial expansion of chaos constructed with integration pointn rules", *J. Mech. Eng. Sci.,* 223(5):1263–1282 (Part C), 2009.

[39] Oberkampf, WL, Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F., Ferson, S.,

"Challenge problems: uncertainty in system response given uncertain parameters", *Reliab. Eng. Syst. Saf.,* 85:11–19, 2004.

[40] Zaman, K., McDonald, M., Mahadevan, S., and Green, L., "Robustness-based design optimization under data uncertainty", *Structural and Multidisciplinary Optimization*, vol. 44, no. 2, pp. 183-197, 2011.

[41] Du, X., Sudjianto, A., Chen, W., "An integrated framework for optimization under uncertainty using inverse reliability strategy", *ASME*, 2004.

[42] Lee I, Choi KK, Du L, Gorsich D, "Dimension reduction method for reliability-based robust design optimization", *Comput. Struct.,* 86(13–14):1550–1562, 2008.

[43] Sundaresan S, Ishii K, Houser DR, "A robust optimization procedure with variations on design variables and constraints". *Eng. Optim.,* 24(2):101–117, 1995

[44] Yu J-C, Ishii K, "Design for robustness based on manufacturing variation patterns" *Trans ASME* 120:196–202, 1998.

[45] Park G-J, Lee T-H, Lee KH, Hwang K-H, "Robust design: an overview" *AIAA J.* 44(1):181–191, 2006.

[46] Haldar, A., Mahadevan, S., "Probability, reliability and statistical methods in engineering design". Wiley, 2000.

[47] Rosenblueth E, "Point estimates for probability moment". *Proc. Natl. Acad. Sci.,* 72(10):3812–3814, 1975.

[48] Zhao Y-G, Ang AH., "System reliability assessment by method of Moments". *J. Struct. Eng.,* 129(10):1341–1349, 2003.

[49] Rahman, S., Xu, H., "A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics". *Probab. Eng. Mech.*, 19:393–408, 2004.

[50] Marler R.T., Arora, J.S., "Survey of multi-objective optimization methods for engineering". *Struct. Multidisc. Optim.*, 26:369– 395, 2004.

[51] Zou, T, Mahadevan, S., "Versatile formulation for multi-objective reliability-based design optimization". *J. Mech. Des.*, 128:1217, 2006.

[52] Mavrotas, G., "Effective implementation of the e-constraint method in multi objective mathematical programming problems", *Appl. Math. Comput.,* 213:455–465, 2009.

[53] Chen, W., Wiecek, M.M., Zhang, J., "Quality utility—a compromise programming approach to robust design", *J. Mech. Des. (ASME)*, 121:179–187, 1999.

[54] Messac, A., Melachrinoudis, E., Sukam, C.P., "Mathematical and pragmatic

perspectives of physical programming", *AIAA J*, 39(5):885–893, 2001.

[55] Agarwal, H., Renaud, J.E, Lee, J., Watson, L.T., "A Unilevel Method for Reliability Based Design optimization", *45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Material Conference*, Palm Spring, California., 2004.

[56] Mohsine, A., Hami, A., "A robust study of reliability-based optimization methods under eigen-frequency" , *Computer Methods in Applied Mechanics and Engineering*, 199 (17–20):1006- 1018, 2010.

[57] Ba-Abbad, M.A., Nikolaidis, E., Kapania, R.K., "New Approach for System Reliability-Based Design Optimization" *AIAA Journal* , 44 (5):1087-1096. doi:10.2514/1.17237 , 2006.

[58] Gano, S.E., Renaud, J.E., Agarwal, H., Tovar, "A Reliability-based design using variable-fidelity optimization", *Structure and Infrastructure Engineering,* 2 (3-4):247-260. doi:10.1080/15732470600590408 ,2006.

[59] Nikolaidis, E., Burdisso, R., "Reliability based optimization: A safety index approach", *Computers & Structures,* 28 (6):781-788. doi:http://dx.doi.org/10.1016/0045-7949(88)90418-X , 1988

[60] Tu, J., Choi, K.K., Park, Y.H., "A New Study on Reliability-Based Design Optimization" *Journal of Mechanical Design,* 121 (4):557-564. doi:10.1115/1.2829499 , 1999.

[61] Du, X., Chen, W., "Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design" *Journal of Mechanical Design,* 126 (2):225-233. doi:10.1115/1.1649968, 2004

[62] Liang, J., Mourelatos, Z.P., Nikolaidis, E., "A Single-Loop Approach for System Reliability- Based Design Optimization" *Journal of Mechanical Design,* 129 (12):1215-1224. doi:10.1115/1.2779884, 2007

[63] Liang, J., Mourelatos, Z.P., Tu, J. "A single-loop method for reliability-based design optimization" *International Journal of Product Development,* 5 (1-2):76-92. doi:doi:10.1504/IJPD.2008.016371, 2008

[64] Shan, S., Wang, G.G., "Reliable design space and complete single-loop reliability-based design optimization" *Reliability Engineering & System Safety*, 93 (8):1218-1230. doi:http://dx.doi.org/10.1016/j.ress.2007.07.006, 2008.

[65] Kira, K., Rendell, L.A., "The feature selection problem: Traditional methods and a new algorithm", in: *Proceedings of Ninth National Conference on Artificial Intelligence*, pp. 129–134, 1992.

[66] Koller, D., Sahami, M., "Toward optimal feature selection", in: *Proceedings of International Conference on Machine Learning*, 1996.

[67] Dash, M., Liu, H., "Feature Selection for Classification", *Intelligent Data Analysis,* 1 131–156. 1997.

[68] Jensen, R., and Shen, Q., "Computational intelligence and feature selection: rough and fuzzy approaches", *Wiley, Hoboken*, NJ, 2008.

[69] Deb, K., "Multi-Objective optimization using evolutionary algorithms", *John Wiley and Sons,* 2001.

[70] Carlos, A., Coello, C., Gary, B., Lamont, D., "Evolutionary Algorithms for Solving Multi-Objective Problems", *Springer*, 2007.

[71] Sharma, S., "Applied Multivariate Techniques", *John Wiley, New York*, 1996.

[72] Saraçli, S., Dogan, N., Dogan, I., "Comparison of hierarchical cluster analysis methods by cophenetic correlation", *Journal of Inequalities and Applications,* Vol. 1, pp.1–8, 2013.

[73] Shannon, C.E., "A mathematical theory of communications", *Bell Syst. Tech. J.*, 27, 379-423, 623-656. 1948.

[74] Chircop, K., amd Zammit-Mangion, D., "On Epsilon-Constraint Based Methods for the Generation of Pareto Frontiers". *Journal of Mechanics Engineering and Automation,* 3. 279-289. 2013.

[75] Szecówka, P.M.,. Szczurek, B.W., "On reliability of neural network sensitivity analysis applied for sensor array optimization", *Sensors and Actuators B: Chemical,* Vol. 157, pp.298–303, 2011.

[76] Bag, A.K., Tudu, B., "Rough Set Based Classification on Electronic Nose Data for Black Tea Application", pp.23–31, 2013.

[77] Doleman, B.J., Lonergan, M.C., Severin, E.J., Vaid, T.S and Lewis, N.S., "Quantitative study of the resolving power of arrays of carbon black-polymer composites in various vapor-sensing tasks", *Analytical Chemistry*, Vol. 70, pp.4177–4190, 1998.

[78] Duda, R., Hart, P.E., Stork, D.G., "Pattern Classification", *John Wiley &Sons*, 2000.

[79] Wijaya, D., Sarno, R., and Zulaika, E., (2016). "Sensor Array Optimization for Mobile Electronic Nose: Wavelet Transform and Filter Based Feature Selection Approach" *International Review on Computers and Software*, Vol. 11, pp 659-671, 2016.