Defence Research and Development Canada
Recherche et développement pour la défense Canada

Data Mining and Fusion in Health Monitoring Applications

Nezih Mrad
Technology Demonstration Section (TDS)
DRDC Atlantic

Zheng Liu
Institute for Research in Construction (IRC)
National Research Council Canada (NRC)

Defence R&D Canada – Atlantic

Technical Memorandum

DRDC Atlantic TM 2011-082

October 2012

Canada

Data Mining and Fusion in Health Monitoring Applications

Nezih Mrad
Technology Demonstration Section (TDS)
DRDC Atlantic

Zheng Liu
Institute for Research in Construction (IRC)
National Research Council Canada (NRC)

Defence R&D Canada – Atlantic

Technical Memorandum
DRDC Atlantic TM 2011-082
October 2012

Principal Author

Original signed by Nezih Mrad

Nezih Mrad

Defence Scientist

Approved by

Original signed by Lex Stuart

Lex Stuart

Head / Technology Demonstration Section

Approved for release by

Original signed by Calvin Hyatt

Calvin Hyatt

DRP/Chair

Abstract

A survey on the state of data mining and fusion technologies and methodologies for structural health monitoring (SHM) is presented in this document. Current research and development efforts are briefly introduced and reviewed. Implementation and application of the diagnostics, prognostics, and health management (DPHM) concepts are also presented, highlighting the significance of data mining and fusion as key components of the concept's architecture. Methodologies and fusion performance metrics are further identified, reviewed and summarized and the potential use of data mining and fusion for SHM and DPHM applications is also discussed. Recommendations on future research and development and on the most promising approaches are also provided.

Résumé

Le document présente une étude sur les technologies et les méthodologies relatives à l'exploration et à la fusion de données pour la surveillance de l'état de structure. On y dresse aussi brièvement la liste des efforts actuels en recherche et en développement. On aborde la mise en oeuvre et l'application des concepts de diagnostics, de pronostics et de gestion de l'état, dans lequel on met en évidence l'importance de l'exploration et de la fusion de données comme composantes clés de l'architecture du concept. Les paramètres de rendement des méthodologies et de la fusion sont recensés, examinés et résumés. On discute de l'emploi possible de l'exploration et de la fusion de données pour les applications de surveillance de l'état de structure et de diagnostics, de pronostics et de gestion de l'état. On formule des recommandations en vue de travaux futurs de recherche et de développement et sur les approches les plus prometteuses.

This page intentionally left blank.

Executive summary

Data Mining and Fusion in Health Monitoring Applications

Nezih Mrad; Zheng Liu; DRDC Atlantic TM 2011-082; Defence R&D Canada – Atlantic; October 2012.

Introduction or background: The philosophy for aircraft maintenance evolves with the emergence of new technologies and methodologies from run-to-failure maintenance and time-based preventive maintenance, to condition-based maintenance (CBM) so that major maintenance expense can be saved. The CBM program only conducts maintenance on the evidence of need, which comes from the information collected though health monitoring. For complex systems like aircraft structures, structural health monitoring (SHM) is a key step to implement a CBM program. In addition, diagnostics and prognostics are two important aspects of the CBM program. Two critical techniques, namely *data fusion* and *data mining*, play a significant role in developing better understanding and interpretation of collected information with an SHM or CBM framework.

Results: This document provides a state-of-the-art review of data fusion and data mining techniques in the realm of aircraft SHM and diagnostics, prognostics, and health management (DPHM). Current research and development work at National Research Council is briefly reviewed. The role of data fusion and data mining in the SHM and/or DPHM systems is identified and described. This report also summarizes the algorithms and methodologies, as well as the fusion performance assessment metrics that have been applied to SHM or DPHM applications. Although most of the techniques themselves are not new and have been used in other fields already, the novel use of these techniques provides a better solution to the specific application. A recommendation for future research and development is given at the end of the report.

Significance: This review on data mining and data fusion methodologies, concepts and techniques introduces the links that exist between SHM, CBM, and DPHM and the significance of the integration of these system components within the overall framework. It is anticipated that this knowledge and understanding will contribute to efforts in the development of a CF CBM strategy.

Future plans: Selected data mining and data fusion approaches will be demonstrated as a component of the current structural health monitoring demonstration activity.

Sommaire

Data Mining and Fusion in Health Monitoring Applications

Nezih Mrad; Zheng Li; DRDC Atlantic TM 2011-082; R & D pour la défense Canada – Atlantique; Octobre 2012.

Introduction ou contexte : La philosophie associée à la maintenance des aéronefs évolue avec l'arrivée de nouvelles technologies et méthodologies, allant de la maintenance visant les défaillances et la maintenance préventive en fonction de la durée, à la maintenance selon l'état (MSE), dans le but de réduire les grandes dépenses en maintenance. Le programme de MSE fonctionne selon les besoins constatés, à partir de l'information recueillie au moyen de la surveillance de l'état. Dans le cas de systèmes complexes comme les structures d'aéronefs, la surveillance de l'état de structure est une étape clé pour la mise en oeuvre d'un programme de MSE. De plus, les diagnostics et les pronostics sont deux aspects importants d'un programme de MSE. Deux techniques essentielles, à savoir l'exploration de données et la fusion de données, jouent un rôle prépondérant dans la compréhension et dans l'interprétation de l'information recueillie dans le cadre des programmes de MSE et de surveillance de l'état de structure.

Résultats : Le document présente une étude de l'état actuel des techniques d'exploration et de fusion de données dans le domaine de la surveillance de l'état de structure et du diagnostic, du pronostic et de la gestion de l'état des aéronefs. Les travaux actuels en recherche et développement au Conseil national de recherches sont brièvement abordés. Le rapport comporte aussi un résumé des algorithmes et des méthodologies , de même que les paramètres d'évaluation du rendement de la fusion des données qui ont servi dans les applications de surveillance de l'état de structure et de diagnostic, de pronostic et de gestion de l'état. Même si la majorité de ces techniques ne sont pas nouvelles en soi, et ont servi dans d'autres domaines, leur utilisation dans de nouveaux domaines offre une meilleure solution pour des applications précises. À la fin du rapport, on formule une recommandation pour des travaux futurs de recherche et de développement.

Importance : Cette étude sur les méthodologies, les concepts et les techniques d'exploration et de fusion de données montre les liens qui existent entre la surveillance de l'état de structure, la maintenance selon l'état (MSE) et le diagnostic, le pronostic et la gestion de l'état et l'importance de l'intégration de ces composantes de système dans le cadre général. On prévoit que ces connaissances contribueront aux efforts de développement d'une stratégie de MSE pour les FAC. Plans futurs : Des approches choisies en matière d'exploration et de fusion de données seront démontrées au titre d'une composante de l'activité actuelle de démonstration de surveillance de l'état de structure.

Plans futurs : Des approches choisies en matière d'exploration et de fusion de données seront démontrées au titre d'une composante de l'activité actuelle de démonstration de surveillance de l'état de structure.

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background

Structural health monitoring (SHM) is defined as a process of implementing a damage identification strategy for aerospace, civil and mechanical engineering infrastructure [1]. The "health" refers to the ability of the structure to continue to perform its intended function in light of inevitable aging and damage accumulation resulting from the operational environments. SHM has recently been recognized as one of the statistical pattern recognition problems. According to the description in [1-2], the SHM process involves the observation of a structure or mechanical system over time with periodic measurements, the extraction of damage-sensitive features from these measurements, and the statistical analysis of these features to determine the current state of system's health. Currently, most structural and mechanical system maintenance is carried out on a time-based mode. The time-based maintenance philosophy is now evolving to a more cost effective condition-based maintenance (CBM) philosophy. Although the major focus of the document is on structural health monitoring, a high level view on the use of data fusion and mining techniques, methodologies and approaches in the context of a much broader concept, i.e. diagnostics, prognostics, and health management (DPHM), is provided.

This document provides a survey on the state of data mining and data fusion techniques and methodologies for structural health monitoring (SHM) and prognostic health management (PHM) applications. It provides a brief overview of some of the research and development efforts. It reflects on the implementation and application of DPHM, highlighting the significance of data mining and fusion as key components of the concept's architecture. Finally, this report presents methodologies and fusion performance metrics with recommendations for future research and development and the most promising approaches.

1.2 Definitions and Concepts

An industry lead, government supported aerospace diagnostics, prognostics health management working group (WG) was established in 2004 [3]. As stated on its official website [4], the primary objective of this working group is to develop and implement a structured approach for continuing consideration of DPHM programs and issues from a Canadian Aerospace sector perspective.

In 2004, the working group developed a technology insight document [3]. This insight document describes the DPHM initiative and technology concepts. The basic definitions and terminology, which are largely derived from the Joint Strike Fighter (JSF) terminology, are given below [5]:

- **Enhanced diagnostics**: the process of determining the state of a component to perform its functions having a high degree of fault detection and fault isolation capability with very low false alarm rate;

- **Prognostics**: the actual material condition assessment which includes predicting and determining the useful life and remaining performance life of components by modeling fault progression;

- **Health management**: the capability to make intelligent, informed, appropriate decisions about maintenance and logistics actions based on diagnostics/prognostics information, available resources and operational demand.

The concepts of data mining and data fusion are also defined in the context of DPHM:

- **Data mining**: machine learning, statistical and soft computing techniques to develop diagnostic and predictive models from data.

- **Data fusion**: the integration of heterogeneous data produced throughout the operation of modern and future aircraft;

The role of data mining and data fusion in the DPHM concept and architecture can be identified from these definitions. Data mining enables the transformation of large amounts of data into useful knowledge which then supports the decision making process. Data fusion is to fuse heterogeneous data in varied formats and at diverse levels. While data fusion can be applied to process sensory data, it can also be used to facilitate decision reasoning through fusing multiple evidences derived from measurements, depending on the application requirements. Therefore, data fusion and mining methods provide learning and reasoning functionalities that can characterize and predict the component or subsystem states from available data [6].

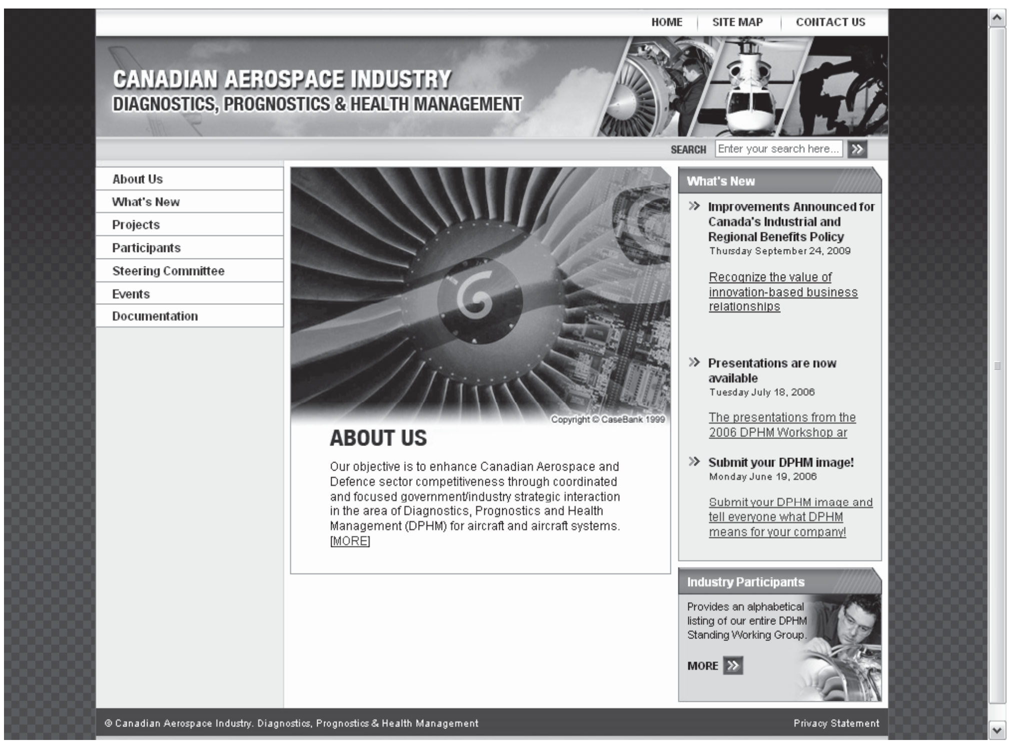


Figure 1: Official website of Canadian aerospace industry DPHM.

1.3 Overview of R&D Efforts

The overview provided in this section focuses on R&D efforts within the National Research Council of Canada (NRC).

1.3.1 Signal-Level Data Fusion

A study by the Institute for Aerospace Research (IAR) of the National Research Council Canada (NRC) investigated the fusion of multi-modal non-destructive inspection (NDI) data for the detection and quantification of hidden corrosion in aircraft lap joints [7]. The NDI techniques considered included multi-frequency eddy current, pulsed eddy current, and enhanced visual inspection. The multi-source data fusion is implemented at the signal level for image enhancement, classification, and quantification. For example, as illustrated in Figure 2, each NDI technique provides an estimation of the material loss through a corresponding calibration procedure. Each measurement is represented by a voltage or gray scale value. The percentage of material loss is estimated from the calibration with such value. To obtain enhanced results, signal-level fusion for the inspection results with different techniques was implemented. Applying wavelets and image pyramid transform, Bayesian inference, Dempster-Shafer evidence theory, and generalized additive model to the multi-modal NDI data, better estimation and characterization of the corroded joints are made [8-10]. The corrosion is characterized by the percentage of material loss by layer. Higher accuracy was achieved when employing data fusion [7, 11-12].

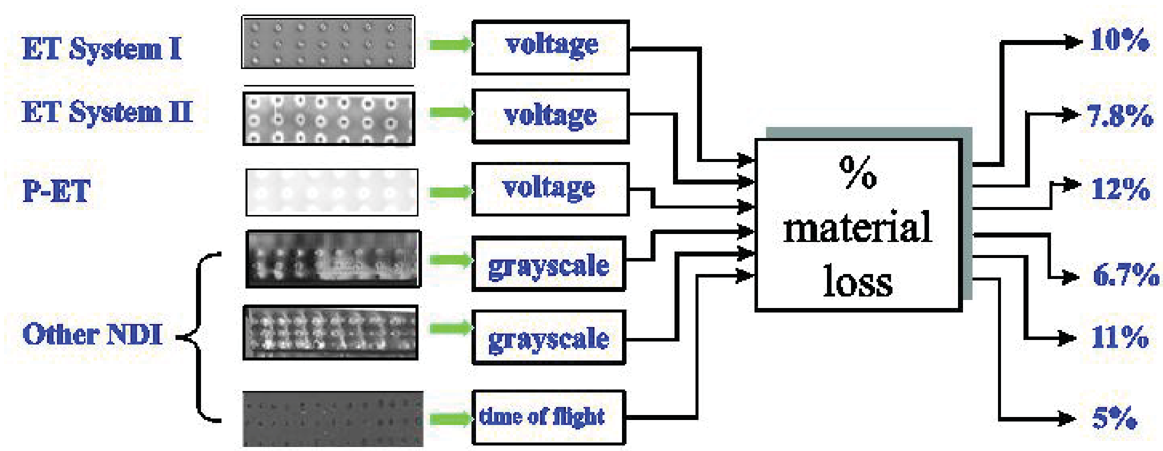


Figure 2: Fusion of multi-modal NDI data for corrosion detection and quantification in aircraft lap joints.

1.3.2 Data Mining

Another study by the Institute for Information Technology (IIT) investigated the use of data mining techniques for the application of aircraft component replacement [13]. Three predictive models were employed: decision tree, instance-based learning, and naive Bayesian learning. These models predicted the need for replacement of various aircraft components based on more than three years of data from a fleet of 34 Airbus A-320. Additionally, the failure of start motor was predicted with rough set theory in [14]. As illustrated in Figure 3 [13], four basic steps are required for the collection and accumulation of representative and valuable data for prediction purposes [15-16]. These are data gathering, data labelling, model building, and model evaluation. It should be highlighted that the building of data mining models requires extensive integrated set of data with specific inputs and outputs.

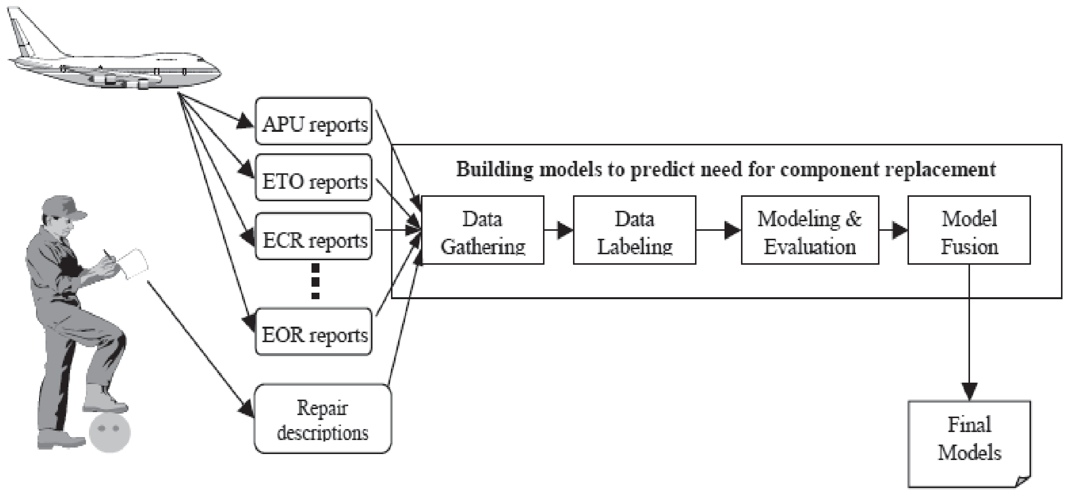


Figure 3: Development of data mining techniques for application of aircraft component replacement.

1.3.3 NRC Collaborative Research

Recently, a collective effort between IAR and IIT, has focused on DPHM applications. An open system architecture condition based maintenance (OSA-CBM) was proposed as shown in Figure 4 [12]. The whole process included data acquisition, manipulation, detection, health assessment, health prediction, and action recommendations.

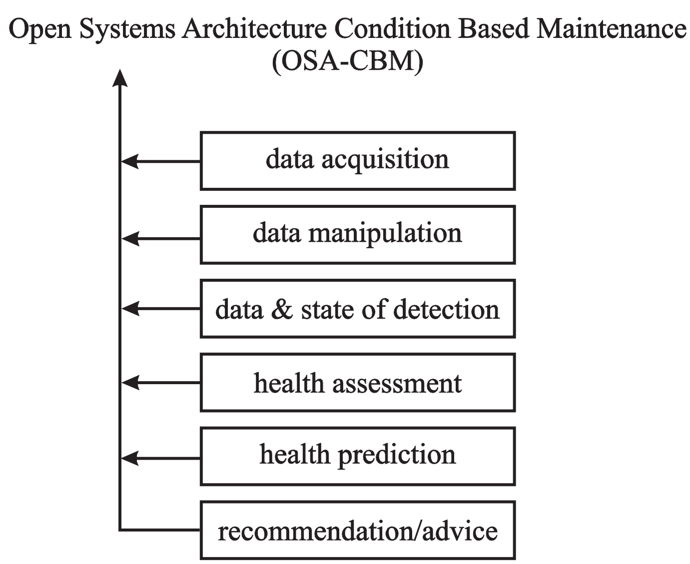


Figure 4: OSA-CBM functional blocks.

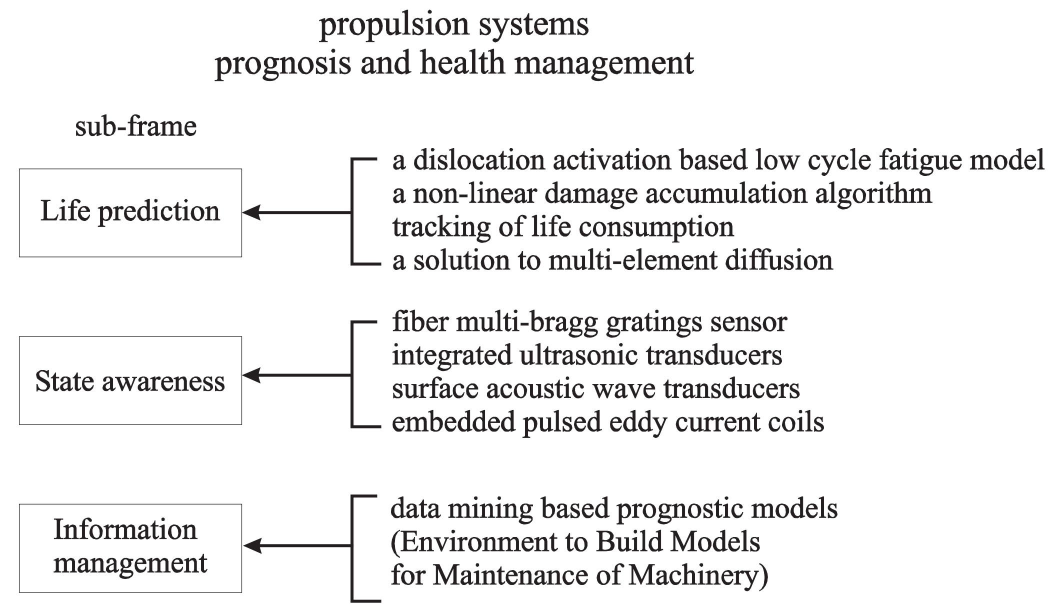


Figure 5: An integrated framework of a Prognostics Health Management system (PHM).

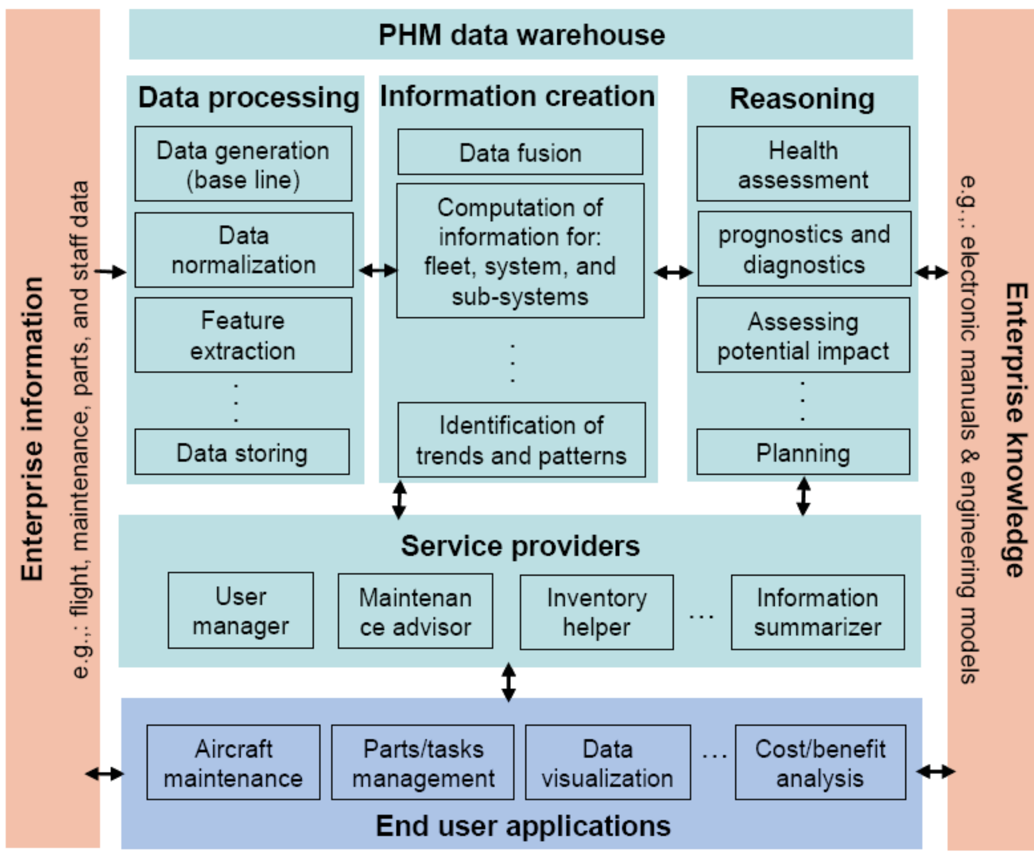


Figure 6: A high-level integrated framework infrastructure.

In addition, an integrated framework of a Prognostics Health Management system (PHM), shown in Figure 5 was developed [17]. The framework that was considered for aircraft propulsion systems consists of three main sub-frames: life prediction, state awareness, and information management. The objective of the sub-frame state awareness is to deliver awareness of component state of health and performance. The information management sub-frame is to build data-driven PHM models for component failure predictions and anomaly detection. These life prediction models will further be fused or combined with the model-based state awareness and life prediction approaches to achieve a more accurate and robust result. An open software platform called EBM3 (Environment to Build Models for Maintenance of Machinery) was developed to incorporate multiple PHM functionalities.

A high-level infrastructure of the integrated framework of a PHM system is given in Figure 6 [17]. To deliver core PHM functionalities, three central modules are identified, i.e. data processing, information creation, and reasoning, where data fusion is identified as a key component in the "information creation" module. The data mining technique could be applied in the "reasoning" module for diagnostics and prognostics.

2 Diagnostics, Prognostics and Health Management

2.1 Concepts

Diagnosis and prognosis are assessment processes for a system's health (past, present, and future) based on observed data and available knowledge [18]. Figure 7 illustrates the functions of diagnosis and prognosis in condition-based maintenance. According to Hess et al. [15,16], prognostics is defined as the capability to provide early detection and isolation of precursor and/or incipient fault condition of a component or sub-element failure , and to manage and predict the progression of the possible failure. This definition includes diagnosis and prognosis for condition-based maintenance (CBM) and decision making. Health management uses health monitoring tools and techniques to detect structural damage, evaluate residual strength and then estimates remaining useful lifetime (RUL) [19].

Two steps are involved in DPHM applications: system development and system implementation [20]. As illustrated in Figure 8, the development process begins from the identification of components or subsystems critical to the performance and reliability of the overall system. Once the components or subsystems are identified, appropriate sensors are selected and instrumented. Physical models of the selected components or subsystems are built and used in model-based diagnosis and prognosis. These models employ algorithms based on physical models. At the implementation step, selected sensors or sensing units collect three types of data: data from healthy and working systems, data from systems with faults, and data from system's transitional failure stage. Features related to the components health state are extracted from the collected data. Then, algorithms for fault and damage detection are applied and the remaining useful life is estimated from the prognostic modules.

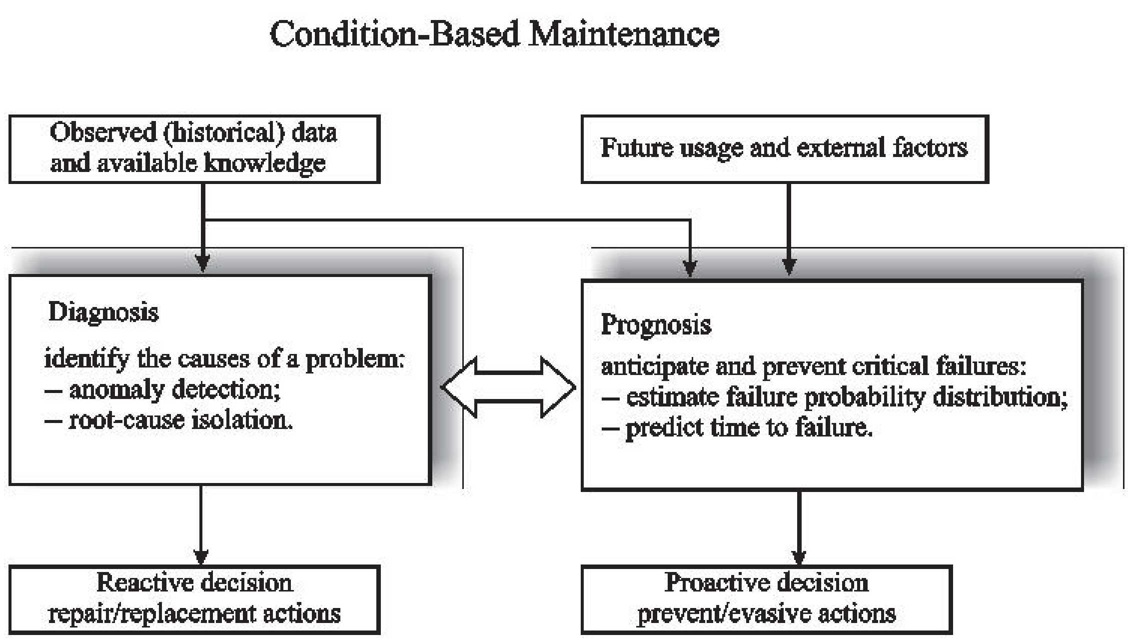


Figure 7: Diagnosis and prognosis in condition-based maintenance.

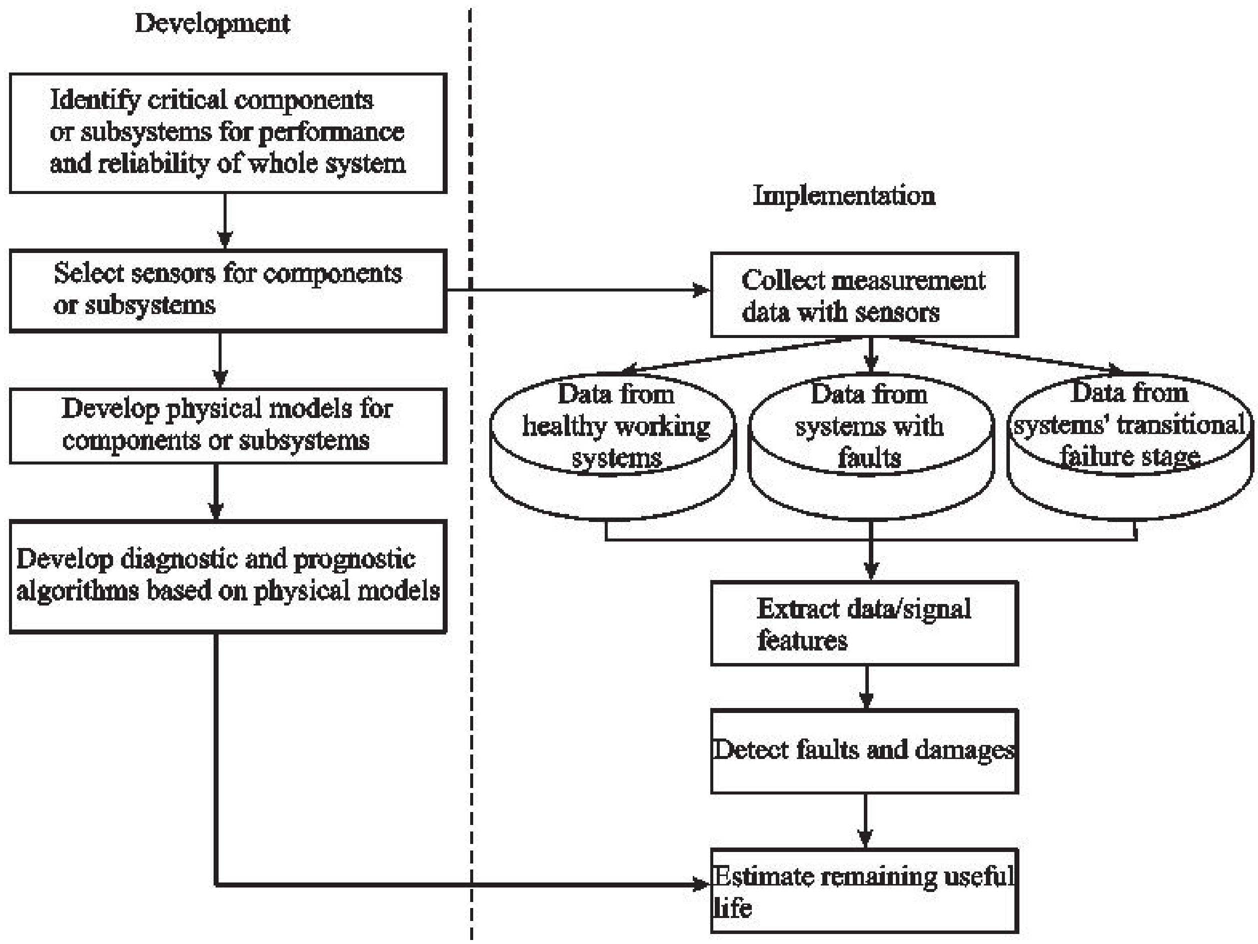


Figure 8: DPHM system development and implementation.

2.2 Implementation

2.2.1 An Integrated Diagnostics and Prognostics Framework

A framework for integrated diagnosis and prognosis system (IDPS) was proposed by Qualtech System Inc. as illustrated in Figure 9 [18]. With this framework, one can detect degradation, anomalies, failures, and causes while being able to assess the health of the system and determine the maintenance requirements.

The IDPS consists of five key components:

1. System configuration editor;
2. Executive;
3. Signal processing module;

4. Diagnostic module;

5. Prognostic module;

6. Database.

The system configuration editor provides an environment for editing system models, diagnostic and prognostic test definitions. The executive is a run-time engine that manages data retrieving from a database or a data acquisition environment. The signal processing module is responsible for feature selection, treatment of reduced/missing data, and training set usage. The diagnosis module provides diagnoses in the presence of multiple and simultaneous failures using a multi-signal modeling methodology. The prognosis module carries out the usage and useful remaining life calculations, while the database is used for local and varied data management.

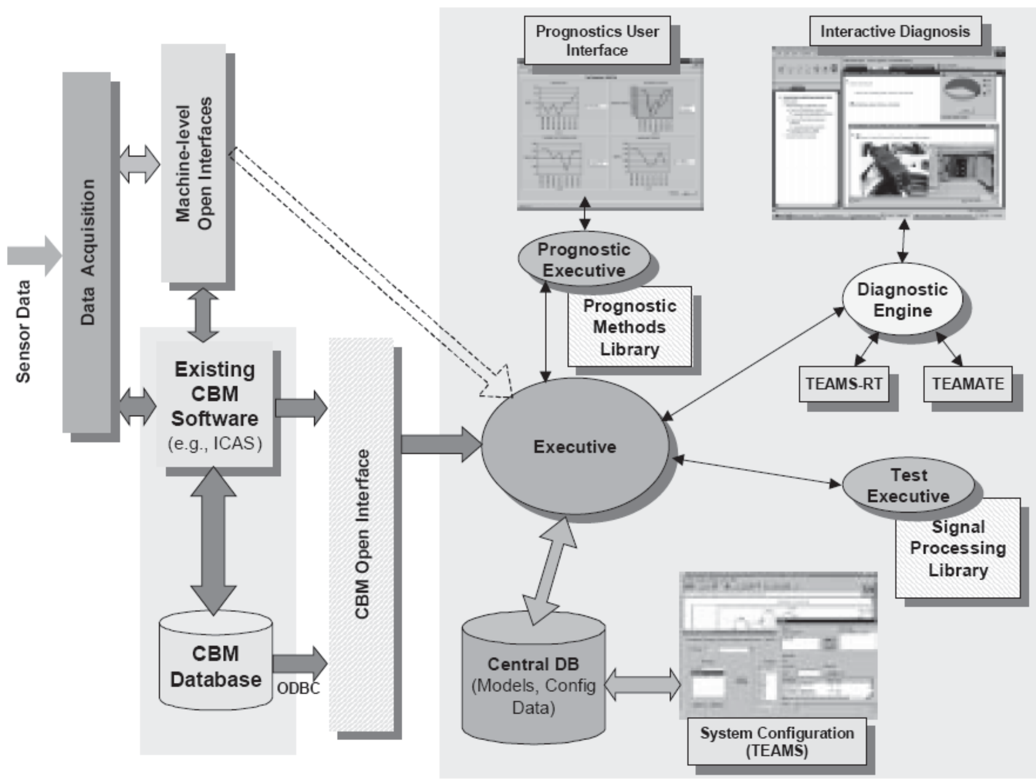


Figure 9: An integrated DPHM system architecture.

2.2.2 Interactive Vehicle Health Management (IVHM) Technology

An IVHM system for air and space transportation systems was proposed in [21] (Figure 10). The goal of this system is to develop validated technologies for automated damage detection, diagnosis, and prognosis [21]. There are two types of diagnostic and monitoring systems, i.e. active sensing and passive sensing, presented in [21]. In the active sensing mode, sensors data are collected and diagnostic information is generated. For example, piezoelectric sensors are active transducers acting for both the generation of controlled diagnostic signals and the collection of measurement data. The passive system is to monitor changes in the environment, such as loads and impacts [22]. This information is sent to other sub-systems for estimating the residual strength and remaining useful lifetime in order to optimize the performance and off-service schedule of the transportation system [21]

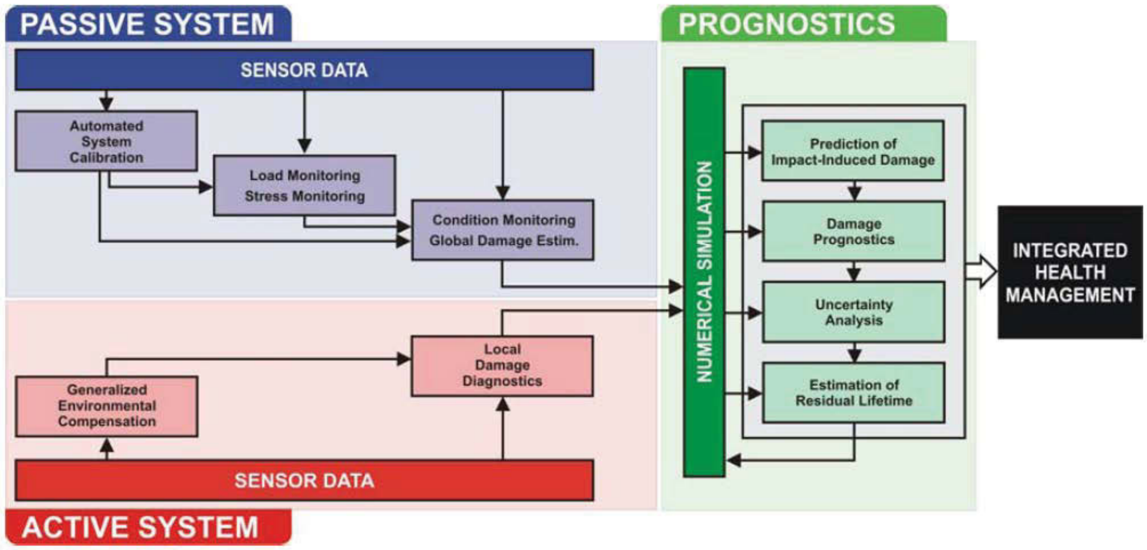


Figure 10: The interactive vehicle health management (IVHM) system.

3 Aircraft DPHM

3.1 Engine Health Management

Prognostic approaches for gas turbine engine health management utilize measured or inferred features as well as different models to predict the condition of the engines' performance and health status. There are seven approaches proposed in [23].

- **Component reliability and usage-based approaches**: In a statistical reliability or usage-based approach, historical failure data or operational usage profile data are used to predict the failure or degradation of a component.

- **Performance trend-based approaches**: In this approach, trend deviations and associated change rates for specific engine features or measurements from normal operating condition are tracked. This approach requires sufficient sensory data and the parametric conditions of a known performance.

- **Data-driven approach**: In this approach, nonlinear network approximators are used to predict future failure based on historical failure data. The data-driven methods include artificial neural networks and fuzzy logic systems.

- **State estimator based approaches**: In this state estimation technique, Kalman filters are used to predict future feature states or systems' behavior through the minimization of error between the model and the measurement.

- **Physics-based modeling approaches**: In physics-based model, the damage as a function of operating conditions, can be calculated and determined.

- **Probability density function for remaining life**: The remaining useful life (RUL) failure probability density function (PDF) is employed to determine the RUL of an engine component. The engine component will be removed from service before attaining a high probability of failure (e.g. a just-in-time point is defined for removal from service that corresponds to a 95% probability that the component has not yet failed).

- **Adaptive prognosis**: In the adaptive prognosis module, current available information is used to update prognosis PDF so that a more accurate prognosis can be established.

The architecture of a distributed prognosis system is illustrated in Figure 11 and as stated in [23], there are many benefits to this type of architecture:

- Optimal computational resource management;

- "Smart subsystem" concept support;

- Multiple faults and damages isolation and assessment;

- Multiple data and information sources capability management;
- Systems degradation capture and localization.

The challenge for the design of a prognostic system remains the ability to fuse measured data and use the results from physics-based models to estimate current and future damage states. The potential for fusing multi-source measurements from C17-T1 flight was investigated [24] and applied to C17-T1 PHM. Positive impact of the data fusion for gas path analysis was observed.

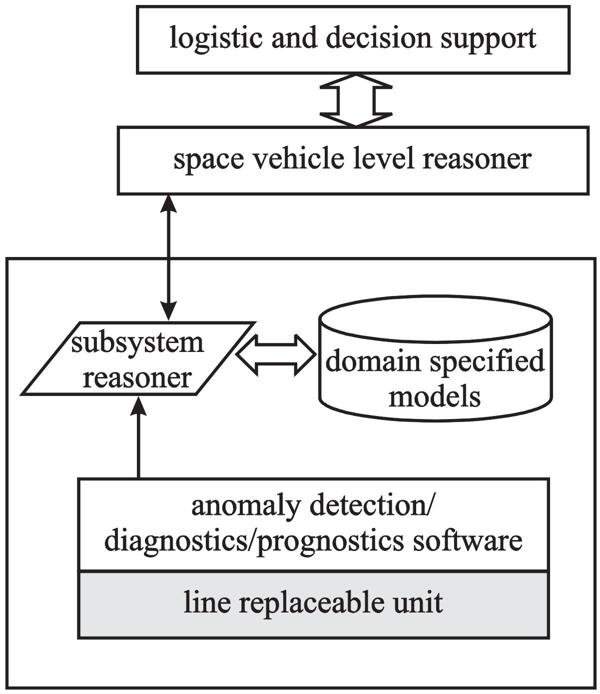


Figure 11: Distributed prognostic system architecture.

3.2 Actuator Health Management

Impact Technologies developed a prognostic and health management (PHM) methodology for aircraft actuator components [25]. This data-driven approach only requires data collected within a flight control system and enables faster algorithm run-times and lower development costs compared with physical modeling. The overall process flow is given in Figure 12 [25].

The flight control data are pre-processed using a “mode detect" algorithm, which recognizes certain operational regimes from the load profiles. The processed data are further extracted for features that are relevant to the current health of the system. Fast Fourier transform and neural networks are employed to extract features from sensor data. A fuzzy logic classification system establishes the relation between extracted features and current health status. The fusion operation combines the operational mode information with the outputs from the classifiers to produce a health state condition [25]. The prognostic reasoner predicts the remaining useful life within specified confidence bounds using the classification and fused information. Kalman filtering is used to predict future health state based on historical health data.

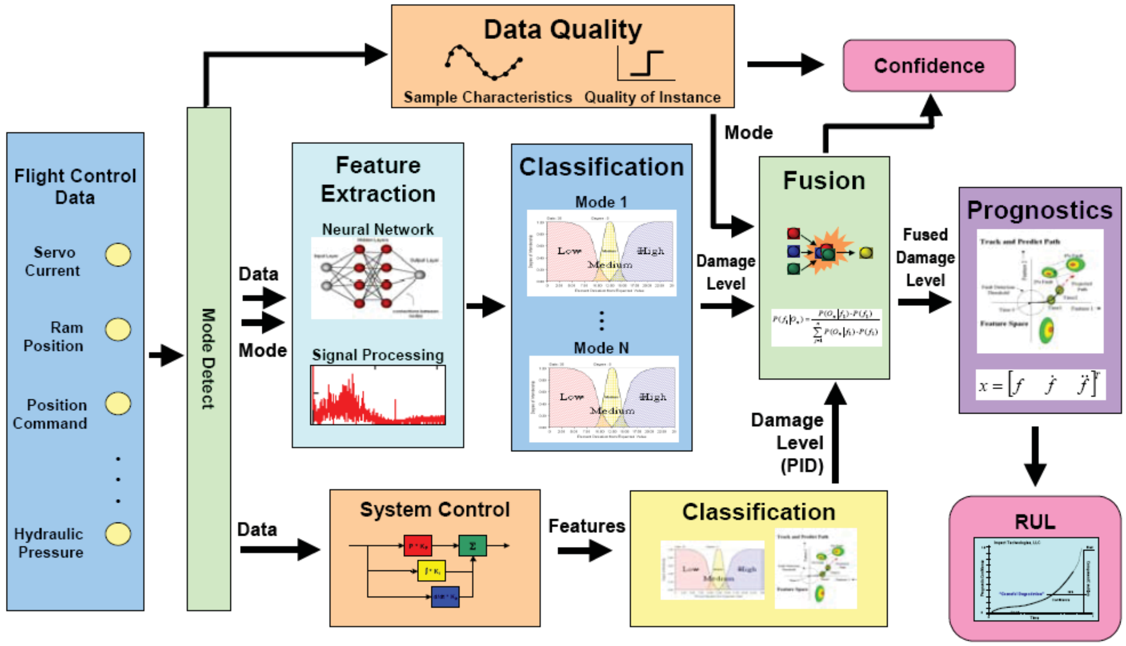


Figure 12: Data-driven methodology for actuator PHM.

This page left intentionally blank

4 Functionalities of Data Fusion and Data Mining

4.1 Concepts of Data fusion and Data Mining

A definition of data fusion, as recommended by the U.S. Department of Defense Joint Directors of Laboratories Data Fusion Subpanel [26], is

> *"data fusion is a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from multiple sources."*

According to [27], the definition of data mining is

> *"data mining is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new data."*

The ultimate goal of data mining is prediction. The techniques used in data mining include: regression, classification, time series, association and sequence analysis, and clustering. It is a combination of three technologies, namely computing power, statistical learning algorithms and tools, and advances in data gathering and management.

4.2 Role of Data fusion and Data Mining

As described in the Canadian Insight Document [28], data mining employs all machine learning, statistical and soft computing techniques to develop data driven diagnostic and predictive models; whereas, data fusion employs techniques and software to integrate heterogeneous data.

4.2.1 Role of Data Fusion

Figure 13 illustrates the role of data fusion within a DPHM process [29]. The fusion of multi-sensory data permits feature extraction and desired signals qualification. Coupled with experience-based information or physical model predictions, these provide optimal diagnostic and prognostics tools.

4.2.1.1 Fusion of information from sensors and models

As described in [30], sensor data are available in two forms: state awareness data and usage data. The ***state awareness sensors*** provide information about the current state of material health from initial indications of defect to crack size estimations. The uncertainties of state awareness sensors include false alarms and measurement errors. ***Usage sensors***, directly or indirectly, provide information on external impacts that may lead to material damage. Data from usage sensors may include information about local stresses and environmental parameters (e.g. such as temperature, humidity, and local chemistry.) The uncertainties of usage sensors include measurement and mapping errors (translation of one type of measurement like accelerations into another data form such as local stresses) [30].

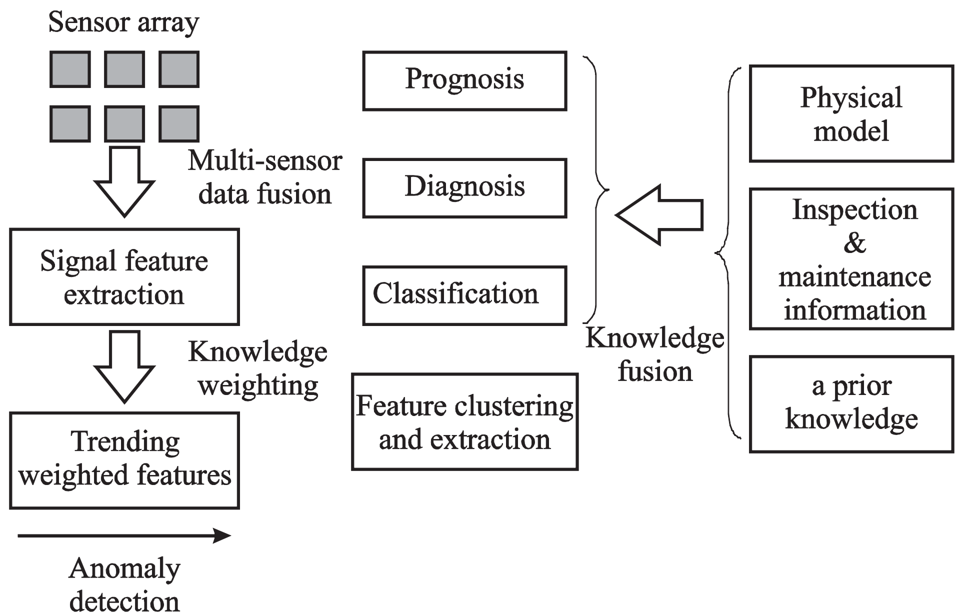


Figure 13: Role of data fusion in DPHM.

Failure models use a variety of input parameters to provide estimates of current state and anticipation of future usage. The input parameters reflect in general material properties and environmental conditions. Due to the limited knowledge of the structure and/or the inadequate representation of its physical characteristics, parameters uncertainties are characterized statistically.

As both models and sensors are imperfect, it is necessary to use the information from both models and sensors to dynamically adjust any predictions. Such predictions reflect health condition at future points, and/or at expected times to reach specified health conditions [30]. The fusion of both sensor and model data reduces the uncertainty associated with output, represented as the probability of failure (POF). A two-stage process of reasoning and prediction fusing sensors and models data is proposed in [30] and is illustrated in Figure 14. In the first stage, sensors are used to detect the presence or absence of a defect, i.e. the current state. The assessment of current state defines the probability density of the time to form a crack of specific size. This assessment is used to update the input parameters to failure models through the Bayesian theorem. The second stage combines updated model predictions and its uncertainties with current state estimates to determine the probability of failure as a function of time and/or usage.

In [31], another fusion scheme based on Kalman filter was proposed to fuse imperfect state information such as environmental measurements with failure models. This type of fusion enables an adaptive prognosis for structural corrosion damages.

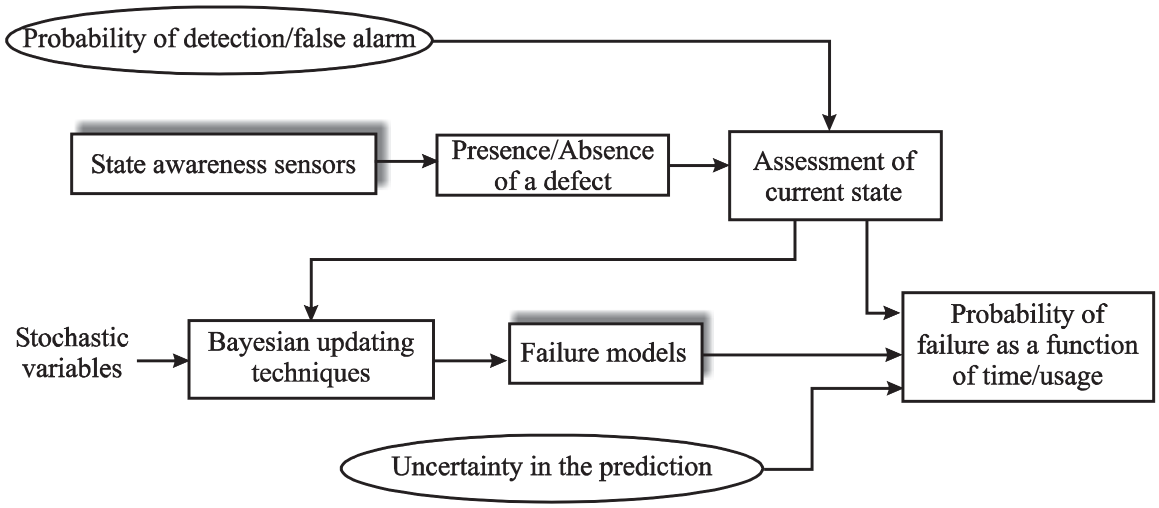


Figure 14: Process of reasoning and prediction by fusing information from sensors and models.

A suit of PHM algorithms were developed for detecting faults in critical bearings associated with aircraft gas turbine engines in [32]. A material-level spall initiation model is used to predict the initiation of a fault. A model that relates the survival rate of the bearing to a stress weighted volume integral developed by Yu and Harris was adopted [32]. However, the modeling predictions have broad confidence bounds due to uncertainties. The health state awareness data from sensors can provide measurements of component condition and can be used to update the modeling assumptions and reduce the uncertainty. Thus, the fusion of sensor data and probabilistic component models can achieve a better decision on the overall health prognostics.

4.2.1.2 Fusion of different prognostic approaches for uncertainty reduction

Both physics-based model and data-driven model can be used to estimate future failure or damage state [33]. The physics-based model requires detailed knowledge about the system, such as material properties and dynamic behavior; whereas, a data-driven model needs sufficient data at known conditions and damage level. Both models implementation possesses pros and cons. The physics-based model relies on the assumption that the fault mode modeled using the specific geometry, material properties, temperature, load, and speed conditions will be similar to the actual fault mode [33]. Any deviations in those parameters will likely result in an error that is amplified over time. The data-driven model assumes the available data sufficiently maps the space and the interpolations/extrapolations from that map can capture the fault rate properly. Therefore, it would be beneficial to fuse the output of both methods which may produce a more accurate and robust result.

In [34], two prognostic models were built to estimate bearing indent damage on outer race. The physics-based model used historic data and estimated future operating conditions (e.g. material properties, geometry, bearing surface interaction, lubrication, and variable operating conditions) to determine future condition by providing a probability density function of the remaining useful life. The data-driven model estimated the spall growth rate based on speed and load. The fusion of these two models is implemented in a prognostic reasoner that employs a combination of damage PDFs, subjective quality assessments, and a kernel-based regression through time [34]

4.2.1.3 Fusion of multiple classifiers for damage location

Piezoelectric accelerometers were applied to locate damage in starboard wing in a Gnat trainer aircraft [35]. Signal features from the accelerometers measurement were extracted. The features selection process was conducted by inspecting the transmissibility functions to find small regions of the frequency range which distinguishes between damage conditions [36]. Then, a multi-layer perceptron (MLP) neural network and a Dempster-Shafer neural network were used to classify the damage locations respectively. The results from these two classifiers were fused with Bayesian and Dempster-Shafer methods to improve the classification rate.

4.2.1.4 Fusion of dynamic and performance analyses

A health monitoring system was developed to diagnose the degradation of aircraft hydraulic pumps [37]. The dynamic analysis of high frequency content of pump pressure and case drain signals provided eight reliable diagnostic features. Each four were used for pump pressure and case drain, respectively. The performance analysis is based on a physics-based approach, which models the performance characteristics of the pump. Fuzzy logic based classification was performed for each analysis approach and Bayesian fusion was applied to fuse the classification results. The procedure for health monitoring of an aircraft hydraulic pump is illustrated in figure 15[37].

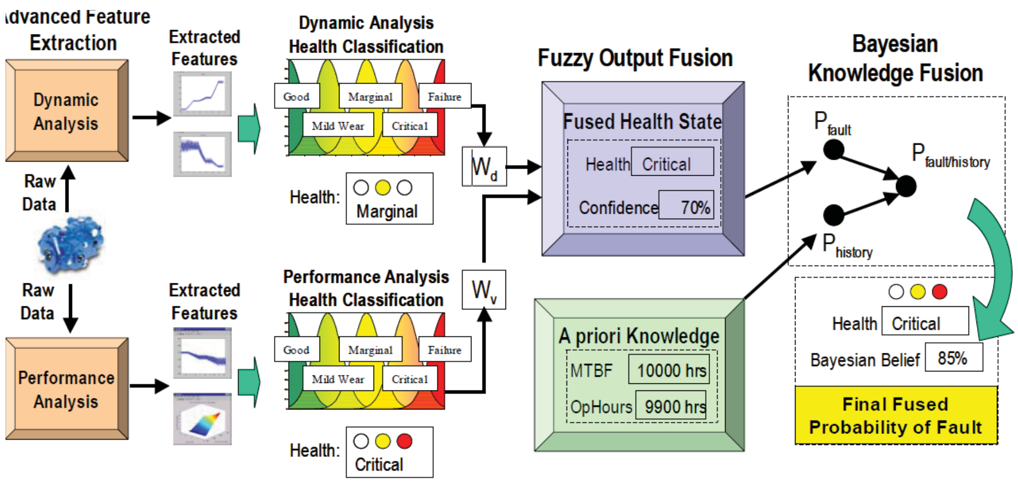


Figure 15: Health monitoring process of an aircraft hydraulic pump.

4.2.2 Role of Data Mining

Data mining searches historical data for unknown patterns. It helps to determine [38]:

- Which component or subsystem should be monitored?
- How often should these components or subsystems be monitored?
- What types of failures occur on a particular component or subsystem?
- What are the warnings of a particular failure?

Data mining methods are also helpful in sensor placement and in determining alert thresholds.

4.2.2.1 Data mining for decision support analyses

Data mining is the process of searching and retrieving useful information from a large data set. It has been applied to decision support in vehicle health management [39]. The process for building and applying data mining models is depicted in Figure 16. The models building step is based on a collected diagnostic, prognostic, and maintenance action data set, where the attributes of past occurrences and corresponding actions are known. In the application step, these models are used to estimate the outcome of current situations. Domain knowledge is usually needed to evaluate the models.

In the modeling step, the raw historical data need to be manipulated to get a set of attribute vectors, which represent a higher level data abstract with complex hybrid information. Such vectors may comprise historical data entries, fault index number, equipment identifiers, test identifiers, and other relevant data. The vectors are fed into the data mining block for the learning process, which generates data mining models. These vectors are used as training examples. The application step applies learned models to new data. The same attribute vectors are created for new data. Decisions are made based on the projected outcome of the current input attributes. It should be mentioned that the quality of the model depends on the quality and span of the data used for training. A detailed discussion can be found in [40].

4.2.2.2 Data mining for system development

Data mining can facilitate the DPHM system development from several aspects [20]. First, data mining techniques can help analyze system degradation mechanisms and identify parts usage, repairs, maintenance, and logistic impacts on component failures. Second, data mining techniques can select appropriate sensors for monitoring based on sensors' reliability, performance records and false alarm rates. The sensor reliability information can be used in the diagnostic and prognostic algorithms, where probabilistic weight may be applied. Third, the models for fault detection, diagnosis, and prognosis can be built with data mining techniques from historical data. The challenge is that there are not enough data, which is statistically significant and representative for the performance of the system.

4.2.2.3 Data mining for system implementation

In the implementation of a DPHM system, data mining can be used to update fault detection and diagnosis algorithms when new data become available. Data mining techniques can also be used to collect relevant data to build models based on system inputs and outputs so that the future usage of the components or subsystems can be predicted. With this information, the prognostic algorithm can estimate the useful remaining life.

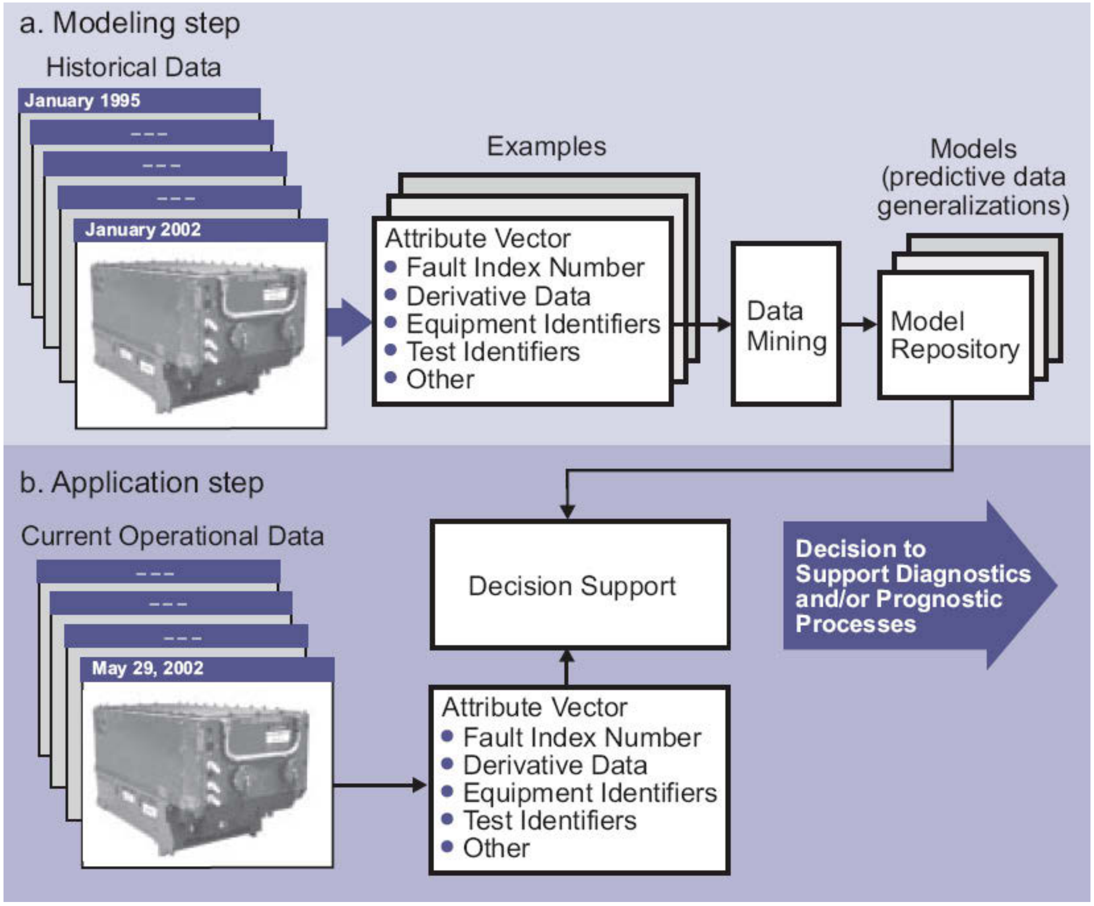


Figure 16: Data mining process for decision support.

4.3 Data Fusion and Data Mining Technologies

Data fusion can be implemented at three levels, i.e. sensor level, feature level, and decision level. At its lowest (sensor) level, the fusion operation combines information from multiple sensors to validate signals and derive features. At its highest level, the fusion operation combines derived features to obtain diagnostic information. Additionally, at this level, the fusion incorporates experience-based information or physical model predictions with signal-based information to facilitate the decision making process. This is illustrated in Figure 17. The implementations for the three-level fusion are different processes. Three approaches to data fusion architectures are presented in Figure 18 [41]. These three approaches correspond to the three levels. The centralized fusion aligns and correlates multi-sensor data in its raw form. The autonomous fusion implements feature extraction before the fusion process. This operation will significantly reduce the dimensionality of the information. The hybrid fusion considers both raw sensory data and extracted features. Therefore, it provides a better solution to practical applications.

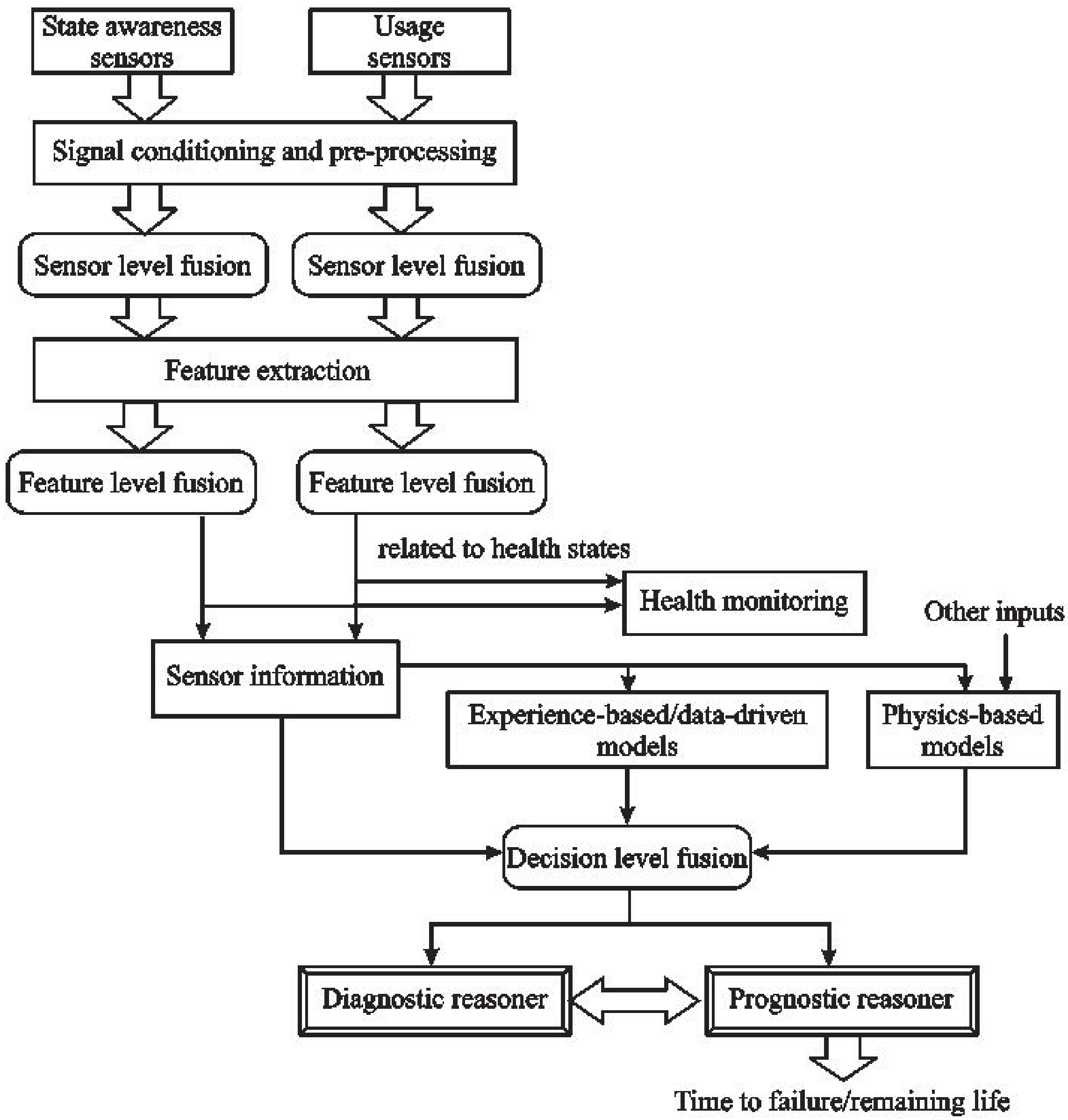
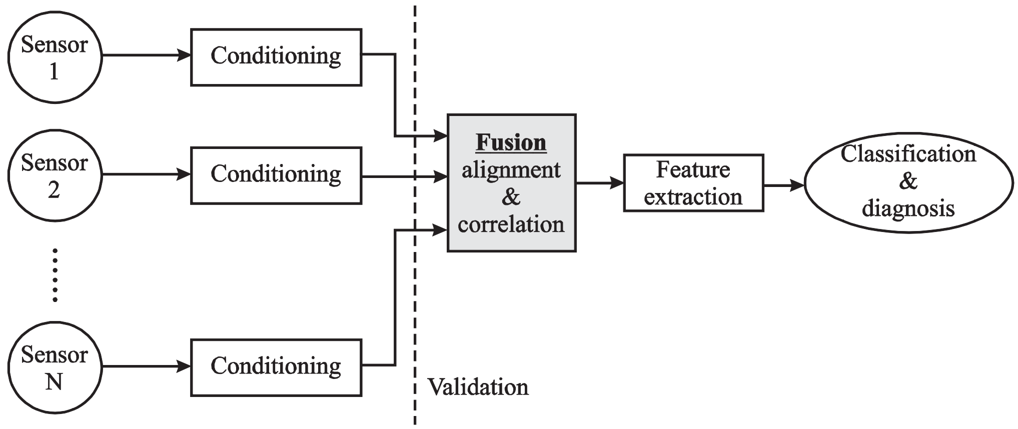
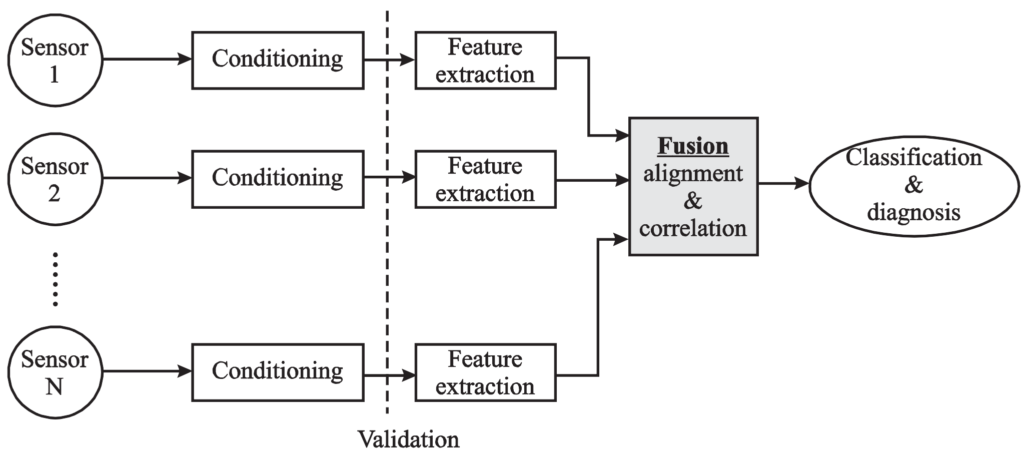


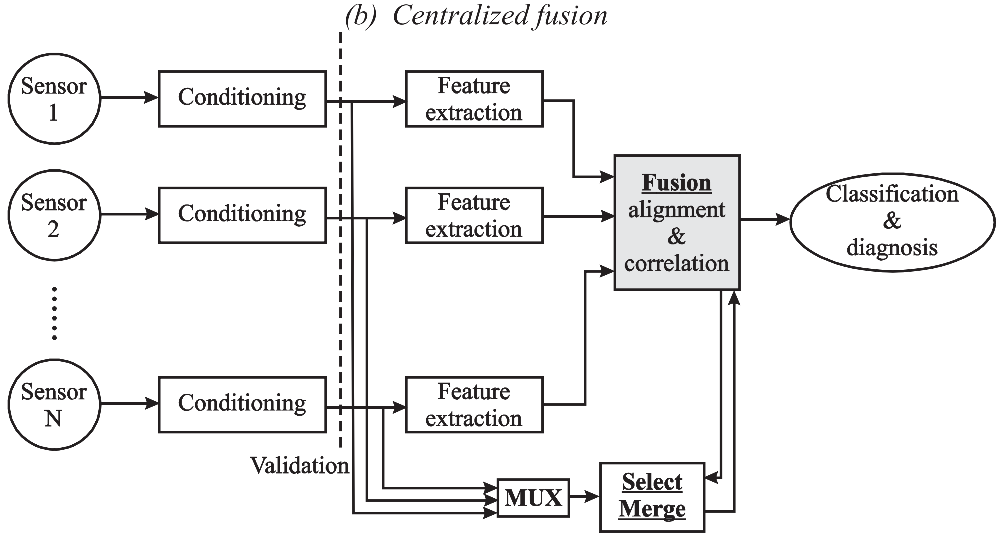
Figure 17: Different level data fusion in a DPHM framework.



(a) Centralized fusion



(b) Centralized fusion



(c) Centralized fusion

Figure 18: Data fusion architectures (MUX: multiplexer).

4.3.1 Data Fusion Algorithms

There is a variety of multi-source data fusion algorithms that have been developed in recent years. A comprehensive review of the fusion methods and descriptions of detailed implementation are available in relevant monographs and literatures [42]. This section discusses only work that has been applied in DPHM applications.

4.3.1.1 Bayesian inference and Dempster-Shafer theory

Bayesian inference provides a mechanism to calculate *a posteriori* probability of a hypothesis being true given support evidence. The hypothesis space for fault types is $\{f_1, f_2, \cdots, f_m\}$, which is mutually exclusive and exhaustive with $\sum_i P(f_i) = 1$. The Bayesian updating is given as:

$$P(f_j \mid O) = \frac{P(O \mid f_j)P(f_j)}{\sum_i P(O \mid f_i)P(f_i)} \tag{4.1}$$

where $P(f_j \mid O)$ is the *a posteriori* probability that fault (f_j) is true given a diagnostic output (O). The probability of getting a diagnostic output (O) for a given fault (f_j) is denoted by $P(O \mid f_j)$ where $P(f_j)$ is the probability of the fault f_j occurring or the *a priori* probability for the fault (f_j). For multiple diagnostic outputs, the probability of fault is expressed as:

$$P(f_j \mid O_1 \cap O_2 \cap \cdots \cap O_n) = \frac{P(f_i)\left[P(O_1 \mid f_j)P(O_2 \mid f_j)\cdots P(O_n \mid f_j)\right]}{\sum_i P(f_i)\left[P(O_1 \mid f_i)P(O_2 \mid f_i)\cdots P(O_n \mid f_i)\right]} \tag{4.2}$$

The probability $P(O_k \mid f_j), (k = 1, 2, \cdots, n)$ can be learned from the available training data. However, the *a priori* probability is not easy to obtain. In other words, the probability of the fault (f_j) occurring is not known in practice. Sometimes, an equal *a priori* probability is assumed and may lead to even worse results.

In the Dempster-Shafer approach, a frame of discernment is constructed for every possible hypothesis. Every hypothesis is assigned a value by a mass function (4.3). An updating process for the mass values $m_i(A)$ and $m_j(B)$ is expressed as:

$$m_{ij}(C) = \frac{\sum_{A \cap B = C} m_i(A)m_j(B)}{1 - \sum_{A \cap B = \varnothing} m_i(A)m_j(B)} \tag{4.3}$$

The output is the belief value, i.e. total probability mass, which provides information on how much a measurement matches the distribution of certain data types. The definition of the mass function is critical for using the Dempster-Shafer approach and depends on the specific application.

4.3.1.2 Disjunctive, conjunctive, and compromise feature fusion

In feature-based fusion, a projection function $\Im$ is used to project a vector of the prior perception $\vec{B}_{prior}$ to a posterior consensus $B_{posterior}$, i.e. [43]

$$\Im : \vec{B}_{prior} \to B_{posterior}, \text{for } \vec{B}_{prior} \in I^n \text{ and } B_{posterior} \in I \tag{4.4}$$

where I is a measure set of the degree of each prior perception, or the degree of the posterior consensus [43]. There are three basic operators for feature-based data fusion: disjunctive, conjunctive, and compromise [44]. If there are two prior perceptions with their degree of x and y, the posterior consensus will be:

- $\Im$ is disjunctive if $\Im(x, y) \geq \max(x, y)$;
- $\Im$ is conjunctive if $\Im(x, y) \leq \min(x, y)$;
- $\Im$ is compromise if $\min(x, y) \leq \Im(x, y) \leq \max(x, y)$.

Considering the case of N sensors, s_j $(j = 1, \cdots, N)$, each sensor possesses a prior perception as to damage, $P(E \mid s_j)$ $(j = 1, \cdots, N)$. There are:

$$\begin{aligned}
&\Im_{dis}(E; s_i, s_2, \cdots, s_N) = \sum_{i=1}^{N} P(E \mid s_i) - \sum_{i=1}^{N-1} \sum_{i<j}^{N} P(E \mid s_i) \cdot P(E \mid s_j) \\
&- \sum_{i=1}^{N-2} \sum_{i<j}^{N-1} \sum_{j<k}^{N} P(E \mid s_i) \cdot PE \mid s_j \cdot P(E \mid s_k) - \cdots \\
&- P(E \mid s_1) \cdot P(E \mid s_2) \cdots P(E \mid s_N)
\end{aligned} \tag{4.5}$$

$$\Im_{conj}(E; s_i, s_2, \cdots, s_N) = P(E \mid s_1) \cdot P(E \mid s_2) \cdots P(E \mid s_N) \tag{4.6}$$

$$\Im_{comp}(E; s_i, s_2, \cdots, s_N) = \frac{1}{N} \sum_{i}^{N} P(E \mid s_i) \tag{4.7}$$

where E can be a variable that characterizes a specific damage parameter, such as location, degree, and orientation [43]. The posterior perception $B_{posterior}$ can be written as:

$$B_{posterior} = \Im(s_1, s_2, \cdots, s_N) \tag{4.8}$$

In practice, it is required to establish the prior perception for all available sensors. There are two steps involved which vary from one application to another:

- Identification of all possible damage locations from individual sensor; and
- Construction of prior probabilities with regard to damage occurrence of each sensor at all locations.

4.3.1.3 Linear combination in the hidden semi-Markov model

A framework based on hidden semi-Markov model (HSMM) was introduced in [45]. The sensor fusion is implemented within this framework by discriminant function analysis. The hidden state at time t is defined by s_t and the observation sequence is defined by O. For a specific component or system, health state can be defined as $H = \{h_1, h_2, \cdots, h_N\}$, where N represents the distinct sequential states for a failure mechanism. If the duration of state i is d_i, the lifetime of the component or system is determined as $T = \sum_{i=1}^{N} d_i$.

For diagnosis, the HSMMs are trained to recognize N different states of a component or system for a given failure mode. The prognosis is implemented by the health-state duration models [45]. Within this hidden semi-Markov model (HSMM) framework, the weights for various sensors are estimated using discriminant function analysis and combined with a linear combination method.

In discriminant function analysis, the weighting procedure is guided by the F values. The F value for a variable indicates its statistical significance in the discrimination between groups. In other words, it is a measure of the extent to which a variable makes a unique contribution to the prediction of group membership. Therefore, the weights for different sensors can be obtained as:

$$w_i = \frac{F_i}{\sum_{i=1}^{N} F_i} \tag{4.9}$$

where w_i denotes the weight for sensor i, F_i denotes the F value for sensor i, and N is the number of sensors. Once the weights for the different sensors are obtained, a linear combination based fusion scheme in HSMMs is developed:

$$Y = \sum_{i=1}^{N} w_i O_i \tag{4.10}$$

where O_i is the measurement from sensor i and Y is the fused result.

4.3.1.4 Fuzzy measures and integrals

The description of fuzzy integral is based on [37]. A fuzzy measure on the set X of criteria is:

$$\mu : P(X) \rightarrow [0,1] \tag{4.11}$$

which satisfies:

$$\mu(\Phi) = 0, \mu(X) = 1 \tag{4.12}$$

$$A \subset B \subset X \text{ implies } \mu(A) \leq \mu(B) \tag{4.13}$$

where $X = \{x_1, \cdots, x_n\}$ is the set of criteria, $P(X)$ is the power set of X, i.e. the set of all subsets of X, $\mu(A)$ represents the weight of importance of the set of criteria A and Φ denotes the empty set.

If a set function $g_\lambda : P(X) \rightarrow [0,1]$ satisfies $g_\lambda(X) = 1$,

- If $A \cap B = \Phi$,
- $g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda \cdot g_\lambda(A) \cdot g_\lambda(B)$
- $\lambda > -1$

then g_λ is a fuzzy measure. The g_λ is determined by

$$g_\lambda = \frac{1}{\lambda}\left[\prod_{i=1}^{n}(1 + \lambda g_i) - 1\right] \tag{4.14}$$

The Choquet fuzzy integral of a function h with respect to μ is defined by

$$C_\mu\left(h(x_1), \cdots, h(x_n)\right) = \sum_{i=1}^{n}\left(h(x_i) - h(x_{i-1})\right)\mu(A_i) \tag{4.15}$$

where $\mu(A)$ is the fuzzy measure representing the importance of the set of criteria A, where $A_i = \{x_i, x_{i+1}, \cdots, x_n\}$.

The fuzzy integral data fusion was applied to the feature level and decision level for machinery fault diagnosis [37] where improved performance is observed.

4.3.2 Data Mining Algorithms

Varied data mining algorithms have been developed and applied to structural health monitoring applications. Table 1 and Table 2 summarize some current usage of these algorithms [46]. As the data mining technique utilizes statistical learning algorithms and tools, which are well documented in literatures like [27], these technical details will not be duplicated in this report.

Table 1: Data mining algorithms applied to fault detection and diagnosis.

Applications	Data mining algorithms	References
Failure detection from sensors	Hidden Markov models (HMM)	[47]
Helicopter fault detection	Ensembles of neural nets	[48]
Inductive monitoring system	Clustering methods	[49]
Rocket propulsion systems	Orca and GritBot	[50]
Space shuttle main engine	Beacon-based exception analysis	[51]
Aircraft avionics diagnosis	Bayesian belief network	[52]
Diagnosis of faults in valves of reciprocating pumps	Support vector machine	[53]

Table 2: Data mining algorithms applied to prognosis.

Applications	Data mining algorithms	References
Structural prognosis	Dynamic wavelet neural networks, reinforcement learning, and genetic algorithm	[54]
Gas turbine engine	Neural nets with rule extractors	[55]
Helicopter gearboxes	Polynomial neural networks	[56]
Batteries	Autoregressive moving average, neural net, and fuzzy logic algorithms	[57]
Complex systems	Bayesian belief net (BBN)	[58]

4.4 Performance Metrics for Data Fusion

The fusion of multi-source data or information is to reduce the uncertainty associated with the sensing and monitoring and improve the accuracy of the produced/expected data or information. The fusion performance evaluation depends on the specific application and the fusion algorithms used. The performance metrics also vary with the fusion level due to the change of requirements. For instance, in the application of defect detection, the sensor level fusion can be assessed by comparing the probability of detection (POD) curve of the fused result with the sensors' POD curves. The benefits of data fusion can also be assessed from system point of view. In [41], the technical value of a diagnostic or detection technology for a particular failure mode is defined as a cost function:

$$\text{TechnicalValue} = P_f * (D * \alpha + I * \beta) - (1 - P_f) * (P_D * \phi - P_f * \theta) \qquad (4.16)$$

where:

P_f : Probability (time-based) of occurrence for a failure mode

D : Overall detection confidence metric score

α : Savings realized by detecting a fault prior to failure

I : Overall isolation confidence metric score

β : Savings realized through automated isolation of a fault

P_D : False positive detection metric score

ϕ : Cost associated with a false positive detection

P_I : False positive isolation metric score

θ : Cost associated with a false positive isolation

The value of a fusion-based diagnostic tool is the summation of the benefits over all the failure modes that it can diagnose less the implementation cost, operation and maintenance cost, and consequential cost of incorrect assessments as expressed:

$$\text{TotalValue} = \sum_{\text{Failure modes}} \text{TechnicalValue}_i - A - O - (1 - P_c) * \delta \tag{4.17}$$

Where,

A : Acquisition and implementation cost

O : Life cycle operation and maintenance cost

P_c : Computer resource requirement score

δ : Cost of a standard computer system

Detailed information about the performance and effectiveness metrics is available in [59]

4.5 Potential Use of Data Mining and Fusion Techniques

The potential use of the data fusion and data mining techniques is summarized in Table 3. It can be seen that the data fusion and data mining techniques are exploited and used in almost every function in SHM or DPHM systems. They provide a flexible and efficient tool for the implementation and integration of a SHM and DPHM system.

Table 3: Potential use of data fusion and mining techniques.

Functionality of SHM and DPHM	Data fusion	Data mining
Identification of critical components or subsystems		√
Sensor selection		√
Sensor validation	√	
Data and signal feature extraction	√	
Fault and damage detection	√	√
Diagnostics	√	√
Prognostics	√	√
Decision making	√	√

5 Summary and Recommendations

This document provided an overview of research and development efforts in the areas of data fusion and data mining techniques and methodologies for diagnostic, prognostics and health management applications, including structural health monitoring. The role of data fusion and data mining techniques in these fields was presented. When the information transits from a low level to a high level, data fusion takes advantage of the heterogeneity with multiple information sources to derive a more accurate and abstract result.

The implementation of data fusion and data mining algorithms is a computational issue, which relies on sensors, available and historic data. The TRL (technology readiness level) of the techniques used depends on how the algorithms are used, e.g. on-line or off-line, onboard or off-board. The technology efficiency also depends on the availability of computational power and resources. The effectiveness or performance of the sensor level fusion is determined by the choice of sensors, which can be complementary to each others. For data mining, the collection of representative historical failure data is important for the success of this technology.

The following recommendations are provided for potential future activities.

1. As described in this report, sensor-level fusion relies on the reliability of measurements, which is typically represented by a ROC (receiver operating characteristics) curve or a POD (probability of detection) curve. The determination of sensor reliability is critical for multi-sensor applications in a DPHM system. Sensor technologies remove some uncertainties associated with human inspection, but introduce new uncertainty associated with sensors. Therefore, a sensor reliability model needs to be developed to understand the uncertainty associated with sensor measurements. Such information can further be used at the sensor data fusion level for reduced prediction uncertainty.

2. Carry out redundancy analysis for sensor failures and sensor anomalies to understand the robustness of the sensing system.

3. Develop a test bed for the evaluation of existing and emerging data mining and data fusion methodologies, algorithms and techniques.

4. Although physics-based models are not easily obtained, the effort on developing such models is still encouraged. It is a source of information for data fusion algorithms and it is a tool that is used to compare data-driven models or/and real-time (or near real-time) sensory data.

5. A software platform named EBM3 (Environment to Build Models for Maintenance of Machinery) was prototyped by the National research Council [60] and was used for executing data mining approaches. An open-architecture platform is needed for the development and implementation of aircraft SHM systems. This platform is not only needed for processing data but also for integrating the whole SHM system. It may encompass all activities for information processing and analysis. Different modules can be implemented and added to this platform, and available for reuse in different applications. It is recommended that the suitability of such a software platform, for example EBM3, should be evaluated.

This page left intentionally blank

References

[1] C.R. Farrar and K. Worden, "An introduction to structural health monitoring," Royal Society of London Transactions Series A, vol. 365, Feb. 2007, pp. 303-315.

[2] C.R. Farrar, S.W. Doebling, and D.A. Nix, "Vibration–based structural damage identification," Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 359, Jan. 2001, pp. 131-149.

[3] Aircraft Systems Diagnostics, Prognostics and Health Management Technology Insight Document, 2004. available at: http://www.dphm-canada.org.

[4] "Aircraft DPHM," 2010, http://www.dphm-canada.org.

[5] A. Hess, G. Calvello, and P. Frith, "Challenges, Issues, and Lessons Learned Chasing the "Big P": Real Predictive Prognostics Part 1," Big Sky, MT, USA: IEEE, 2005, pp. 1-10.

[6] D. Raheja, J. Llinas, R. Nagi, and C. Romanowski, "Data fusion/data mining-based architecture for condition-based maintenance," International Journal of Production Research, vol. 44, Jul. 2006, pp. 2869-2887.

[7] Z. Liu, D.S. Forsyth, P. Ramuhalli, and A. Fahr, "A Data Fusion Framework for Multiple Nondestructive Inspection Images," C.H. Chen, ed., World Scientific Publishing Co. Pte. Ltd., 2007, pp. 385-402.

[8] Z. Liu, D.S. Forsyth, M.S. Safizadeh, M. Genest, A. Fahr, and A. Marincak, "Fusion of Visual and Eddy Current Inspection Results for the Evaluation of Corrosion Damage in Aircraft Lap Joints," SPIE Proceedings 5768, San Diego, CA, USA: 2005.

[9] Z. Liu, D.S. Forsyth, M.S. Safizadeh, B.A. Lepine, and A. Fahr, "Quantitative Interpretation of Multi-frequency Eddy Current Data by Data Fusion Approaches," Proceedings of SPIE, San Diego, CA, USA.: 2003, pp. 39-47.

[10] Z. Liu, P. Ramuhalli, S. Safizadeh, and D.S. Forsyth, "Combining multiple nondestructive inspection images with a generalized additive model," Measurement Science and Technology, vol. 19, Aug. 2008, p. 85701.

[11] Z. Liu, D.S. Forsyth, M.S. Safizadeh, and A. Fahr, "A Data Fusion Scheme for Quantitative Image Analysis by Using Locally Weighted Regression and Dempster-Shafer Theory," IEEE Transactions on Instrumentation and Measurement, vol. 57, 2008, pp. 2554-2560.

[12] Z. Liu, A. Fahr, and N. Mrad, "Application of Dempster-Shafer Theory for the Fusion of Lap Joints Inspection Data," A.A. Mufti, A.L. Gyekenyesi, and P.J. Shull, eds., 2006, pp. 337-345.

[13] S. Letourneau, F. Famili, and S. Matwin, "Data Mining to Predict Aircraft Component Replacement," IEEE Intelligent Systems, vol. 14, 1999, pp. 59-66.

[14] J.M. Pena, S. Létourneau, and F. Famili, “Application of Rough Sets Algorithms to Prediction of Aircraft Component Failure,” IDA ’99: Proceedings of the Third International Symposium on Advances in Intelligent Data Analysis, London, UK: Springer-Verlag, 1999, pp. 473-486.

[15] A. Hess, G. Calvello, and P. Frith, “Challenges, Issues, and Lessons Learned Chasing the "Big P": Real Predictive Prognostics Part 1,” 2005 IEEE Aerospace Conference, IEEE, 2005, pp. 1-10.

[16] A. Hess, G. Calvello, P. Frith, S.J. Engel, and D. Hoitsma, “Challenges, Issues, and Lessons Learned Chasing the "Big P": Real Predictive Prognostics Part 2,” 2006 IEEE Aerospace Conference, IEEE, 2006, pp. 1-19.

[17] J. Bird, X. Wu, P. Patnaik, A. Dadouche, S. Letourneau, and N. Mrad, “A Framework of Prognosis and Health Management - A Multidisciplinary Approach,” Proceedings of GT2007, Montreal, Canada: 2007, pp. 1-10.

[18] A. Mathur, K.F. Cavanaugh, K.R. Pattipati, P.K. Willett, and T.R. Galie, “Reasoning and modeling systems in diagnosis and prognosis,” SPIE Conference Series, P. K. Willett & T. Kirubarajan, eds., 2001, pp. 194-203.

[19] I. Mueller, C. Larrosa, S. Roy, and F.-K. Chang, “An Integrated Diagnostic and Prognostic SHM Technology for Structural Health Management,” Proceedings of Structural Health Monitoring, 2009, pp.1-15.

[20] K. Reichard, E. Crow, and L. Weiss, “Applications of Data Mining in Automated ISHM and Control for Complex Engineering Systems,” Proceedings of the first international forum on integrated system health engineering and management in aerospace, 2005, pp. 1-13.

[21] I. Mueller, C. Larrosa, S. Roy, A. Mittal, K. Lonkar, and F.-K. Chang, “An Integrated Health Management and Prognostic Technology for Composite Airframe Structures,” Annual Conference of the Prognostics and Health Management Society, 2009, pp. 1-15.

[22] R. Seydel and F.-K. Chang, “Impact identification of stiffened composite panels: I. System development,” Smart Materials and Structures, vol. 10, 2001, pp. 354-369.

[23] M.J. Roemer, C.S. Byington, G.J. Kacprzynski, and G. Vachtsevanos, “An Overview of Selected Prognostic Technologies with Application,” Proceedings of GT2006, Barcelona, Spain: 2006, pp. 1-9.

[24] A. Volponi, “Data Fusion for Enhanced Aircraft Engine Prognostics and Health Management,” Technical Report, NASA/CR-2005-214055, Dec. 2005.

[25] C.S. Byington, M. Watson, and D. Edwards, “Data-driven neural network methodology to remaining life predictions for aircraft actuator components,” 2004 IEEE Aerospace Conference Proceedings, IEEE, 2004, pp. 3581-3589.

[26] L.A. Klein, Sensor and Data Fusion Concepts and Applications, Bellingham, WA, USA: Society of Photo-Optical Instrumentation Engineers (SPIE), 1999.

[27] T. Hill and P. Lewicki, Statistics: Methods and Applications, StatSoft, Inc., 2006.

[28] Aircraft Systems Diagnostics, Prognostics and Health Management Technology Insight Document, Ottawa, Ontario, Canada: Industry Canada, 2004.

[29] M.J. Roemer, G.J. Kacprzynski, and R.F. Orsagh, "Assessment of data and knowledge fusion strategies for prognostics and health management," Proceedings of the 2001 IEEE Aerospace Conference, Big Sky, Montana, March, vol. 6, 2001, p. 2979-2988.

[30] J.M. Papazian, E.L. Anagnostou, S.J. Engel, D. Hoitsma, J. Madsen, R.P. Silberstein, G. Welsh, and J.B. Whiteside, "A structural integrity prognosis system," Engineering Fracture Mechanics, vol. 76, 2009, pp. 620-632.

[31] G.J. Kacprzynski and D.S. Muench, "Sensor/Model Fusion for Adaptive Prognosis of Structural Corrosion Damage," 2006, pp. 1-5.

[32] M.J. Roemer and C.S. Byington, "Prognostics and health management software for gas turbine engine bearings," Proceedings of GT2007, 2007, pp. 1-8.

[33] K. Goebel, N. Eklund, and P. Bonanni, "Fusing Competing Prediction Algorithms for Prognostics," 2006 IEEE Aerospace Conference, IEEE, 2006, pp. 1-10.

[34] K. Goebel, N. Eklund, and P. Bonanni, Prognostic Fusion for Uncertainty Reduction, 2007.

[35] G. Manson, "EXPERIMENTAL VALIDATION OF A STRUCTURAL HEALTH MONITORING METHODOLOGY: PART III. DAMAGE LOCATION ON AN AIRCRAFT WING," Journal of Sound and Vibration, vol. 259, Jan. 2003, pp. 365-385.

[36] K. Worden, G. Manson, and T. Denœux, "An evidence-based approach to damage location on an aircraft structure," Mechanical Systems and Signal Processing, vol. 23, 2009, pp. 1792-1804.

[37] S. Amin, C.B.P. E, and M. Watson, "Fuzzy Inference and Fusion for Health State Diagnosis of Hydraulic Pumps and Motors," Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society, 2005, pp. 1-6.

[38] D. Raheja, J. Llinas, R. Nagi, and C. Romanowski, "Data fusion/data mining-based architecture for condition-based maintenance," International Journal of Production Research, vol. 44, Jul. 2006, pp. 2869-2887.

[39] J.A. Castrigno, S.J. Engel, and B.J. Gilmartin, "Vehicle Health Management : Architectures and Technologies," Technology Review, 2006, pp. 29-46.

[40] K. Worden, C.R. Farrar, G. Manson, and G. Park, "The fundamental axioms of structural health monitoring," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 463, 2007, pp. 1639-1664.

[41] M.J. Roemer, G.J. Kacprzynski, and R.F. Orsagh, "Assessment of data and knowledge fusion strategies for prognostics and health management," Aerospace Conference, 2001, IEEE Proceedings, 2001, pp. 2979-2988.

[42] D.L. Hall, Mathematical Techniques in Multisensor Data Fusion, Norwood, MA, USA: Artech House, Inc., 1992.

[43] Z. Su, X. Wang, L. Cheng, L. Yu, and Z. Chen, "On Selection of Data Fusion Schemes for Structural Damage Evaluation," Structural Health Monitoring, vol. 8, May. 2009, pp. 223-241.

[44] I. Bloch, "Information combination operators for data fusion: a comparative review with classification," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 26, 1996, pp. 52-67.

[45] M. Dong and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," European Journal of Operational Research, vol. 178, May. 2007, pp. 858-878.

[46] M.A. Schwabacher, "A Survey of Data-Driven Prognostics," AIAA InfoTech Aerospace, 2005.

[47] A.N. Srivastava, Ph, "Discovering System Health Anomalies Using Data Mining Techniques," Proceedings of the Joint Army Navy NASA Air Force Conference on Propulsion, 2005.

[48] N. Oza, K. Tumer, I. Tumer, and E. Huff, "Classification of Aircraft Maneuvers for Fault Detection," Multiple Classifier Systems, T. Windeatt and F. Roli, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, p. 160.

[49] D.L. Iverson, "Inductive system health monitoring," In Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI04), Las Vegas, 2004.

[50] M. Schwabacher, "Machine Learning for Rocket Propulsion Health Monitoring," Proceedings of the SAE World Aerospace Congress, Dallas TX, USA: 2005.

[51] H. Park, R. Mackey, M. James, M. Zak, M. Kynard, J. Sebghati, and W. Greene, "Analysis of space shuttle main engine data using beacon-based exception analysis for multi-missions," Proceedings of the IEEE Aerospace Conference, New York, USA: 2002, pp. 2835-2844.

[52] C.S. Byington, P.W. Kalgren, R. Johns, and R.J. Beers, "Embedded Diagnostic/Prognostic Reasoning and Information Continuity for Improved Avionics Maintenance," Proceedings of AUTOTESTCON, 2003, pp. 320-329.

[53] F. He and W. Shi, "WPT-SVMs based approach for fault detection of valves in reciprocating pumps," Proceedings of the 2002 American Control Conference, American Automatic Control Council, 2002, pp. 4566-4570.

[54] P. Wang and G. Vachtsevanos, “Fault prognostics using dynamic wavelet neural networks,” Artif. Intell. Eng. Des. Anal. Manuf., vol. 15, 2001, pp. 349-365.

[55] T. Brotherton, G. Jahns, J. Jacobs, and D. Wroblewski, “Prognosis of faults in gas turbine engines,” Aerospace Conference Proceedings, 2000 IEEE, 2000, pp. 163 -171 vol.6.

[56] B.E. Parker, T.M. Nigro, M.P. Carley, R.L. Barron, D.G. Ward, H.V. Poor, D. Rock, and T.A. DuBois, “Helicopter gearbox diagnostics and prognostics using vibration signature analysis,” Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, S. K. Rogers, eds., 1993, pp. 531-542.

[57] L. Shimanek, “Battery Prognostics.”, empfasis, December 2003, pp.1-2.

[58] K.W. Przytula and A. Choi, “Reasoning framework for diagnosis and prognosis,” Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA: 2007.

[59] R.F. Orsagh, M.J. Roemer, C.J. Savage, and M. Lebold, “Development of performance and effectiveness metrics for gas turbine diagnostic technologies,” Big Sky, MT, USA: 2002, pp. 2825-2833.

[60] J. Bird, X. Wu, P. Patnaik, A. Dadouche, S. Letourneau, and N. Mrad, “A Framework of Prognosis and Health Management - A Multidisciplinary Approach,” Montreal, Canada: 2007.

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Atlantic 9 Grove Street P.O. Box 1012 Dartmouth, Nova Scotia B2Y 3Z7	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC December 2013

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Data Mining and Fusion in Health Monitoring Applications

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Mrad, N; Liu, Z.

5. DATE OF PUBLICATION (Month and year of publication of document.) October 2012	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 52	6b. NO. OF REFS (Total cited in document.) 60

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Atlantic
9 Grove Street
P.O. Box 1012
Dartmouth, Nova Scotia B2Y 3Z7

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.) 13PL	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Atlantic TM 2011-082	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

A survey on the state of data mining and fusion technologies and methodologies for structural health monitoring (SHM) is presented in this document. Current research and development efforts are briefly introduced and reviewed. Implementation and application of the diagnostics, prognostics, and health management (DPHM) concepts are also presented, highlighting the significance of data mining and fusion as key components of the concept's architecture. Methodologies and fusion performance metrics are further identified, reviewed and summarized and the potential use of data mining and fusion for SHM and DPHM applications is also discussed. Recommendations on future research and development and on most promising approaches are also provided.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Data fusion; data mining; SHM; CBM; PHM; DPHM; aircraft; signal processing.

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca