Design and Mining of Health Information Systems for Process and Patient Care Improvement

by

Kun Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Jing Li, Chair
Teresa Wu
Rong Pan
Christine M Zwart

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

In healthcare facilities, health information systems (HISs) are used to serve different purposes. The radiology department adopts multiple HISs in managing their operations and patient care. In general, the HISs that touch radiology fall into two categories: tracking HISs and archive HISs. Electronic Health Records (EHR) is a typical tracking HIS, which tracks the care each patient receives at multiple encounters and facilities. Archive HISs are typically specialized databases to store large-size data collected as part of the patient care. A typical example of an archive HIS is the Picture Archive and Communication System (PACS), which provides economical storage and convenient access to diagnostic images from multiple modalities. How to integrate such HISs and best utilize their data remains a challenging problem due to the disparity of HISs as well as high-dimensionality and heterogeneity of the data. My PhD dissertation research includes three inter-connected and integrated topics and focuses on designing integrated HISs and further developing statistical models and machine learning algorithms for process and patient care improvement.

Topic 1: Design of super-HIS and tracking of quality of care (QoC). My research developed an information technology that integrates multiple HISs in radiology, and proposed QoC metrics defined upon the data that measure various dimensions of care. The DDD assisted the clinical practices and enabled an effective intervention for reducing lengthy radiologist turnaround times for patients.

Topic 2: Monitoring and change detection of QoC data streams for process improvement. With the super-HIS in place, high-dimensional data streams of QoC metrics are generated. I developed a statistical model for monitoring high-dimensional data streams that integrated Singular Vector Decomposition (SVD) and

process control. The algorithm was applied to QoC metrics data, and additionally extended to another application of monitoring traffic data in communication networks.

Topic 3: Deep transfer learning of archive HIS data for computer-aided diagnosis (CAD). The novelty of the CAD system is the development of a deep transfer learning algorithm that combines the ideas of transfer learning and multi-modality image integration under the deep learning framework. Our system achieved high accuracy in breast cancer diagnosis compared with conventional machine learning algorithms.

ACKNOWLEDGMENTS

To begin with, I would like to thank my advisor, Jing Li, who taught me the way of work independently and who encouraged and support me during my research and dissertation. Her attitude of life and research profoundly influenced me during the last several years and also would help me through lifelong. Gratitude to all members of my committee: Teresa Wu, Rong Pan, and Christine Zwart. A special thank you to my collaborators at Mayo Clinic, Ross Mitchell and Clinton Wellnitz, both help me step deeper into the healthcare field. A warm acknowledgment to Bhavika Patel and Bin Zheng for their valuable interactions and feedback.

Thank you to all members of AMIIL, past and present: Can Cui, Na Zou, Bing Si, Shuluo Ning, Nathan Gaw, Fei Gao, Min Zhang, Yinlin Fu, Xiaonan Liu, Hyunsoo Yoon, Lujia Wang, Congzhe Su, and Yanzhe Xu. A special thank to Nathan Gaw, we went through the tough first year together. Thank you all for the help and support to Cheng Niu, Randy Zhong during my internship in Palo Alto.

More personally, thank you to roommates, particularly Chao Li, Di Yang and Shuluo Ning, you listened, helped, and empathized. Very specially, I'd like to appreciate my parents and my wife for unconditionally supports all days and nights, all winters and summers.

TABLE OF CONTENTS

Page

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Background

The radiology department is one of the busiest departments in a healthcare facility, as it serves the need of many other departments for providing imaging service and radiologic reports. In radiology, a number of Healthcare Information Systems (HISs) are used to serve different purposes, such as recording patient information and care process and storing scanned images for disease diagnosis and evaluation. Generally speaking, the HISs that touch Radiology fall into one of two categories: tracking HISs and archive HISs. Electronic Health Records (EHR) is a typically tracking HIS, which tracks the care each patient receives at multiple encounters and facilities, such as progress notes, medications, immunizations, vital signs, laboratory and radiologic reports. Radiology Information Systems (RIS) is another tracking HIS specifically designed for radiology, which tracks patients, exams, result distribution, and procedure billing. Different from tracking HISs, archive HISs are typically specialized databases to store large-size data collected as part of the patient care. A typical example of an archive HIS is the Picture Archive and Communication System (PACS), which provides economical storage and convenient access to images from multiple modalities. PACS stores both image data as well as image related meta-data information.

The availability of multiple HISs in radiology provides an unprecedented opportunity for improving patient care. On the other hand, how to integrate these systems and best utilize the data remains a challenging problem. In my dissertation, I investigate three topics related the design/integration and data-mining of HISs for process and patient care improvement. The first two topics are related to tracking HISs and the third topic is related to archive HISs especially radiologic images stored in PACS.

1.2 Summary of Research Topics and State of the Art

Topic 1: **Integration of Multiple Health Information Systems for Quality Improvement of Radiologic Care.** As previously introduced, several HISs are commonly used in radiology, including EHR, RIS, and PACS. Each HIS records partial and complementary information about the radiologic care process. Depending on the institution, the HISs that touch radiology can be distinct, disparate, and with different database formats and meta information semantics. Due to these practical challenges, we note no reported research on integrating multiple HISs to allow for an end-to-end tracking of the care patients receive in the radiology department. Therefore, the current Quality of Care (QoC) research is limited as it can only utilize data from a siloed HIS for partial workflow analysis. A comprehensive assessment of the QoC in radiology requires multiple HISs be integrated such that various key QoC metrics over the end-to-end workflow can be extracted. Driven by this, we developed a novel technology called Department Data Depot (DDD) that integrates multiple HISs in radiology. We proposed nine QoC metrics defined upon the data from DDD that measure various dimensions of care quality such as timeliness, efficiency, patient satisfaction, and workload distribution. To demonstrate the clinical utility of DDD, we developed and deployed a web application system, the Radiology Quality Dashboard (RQD), at Mayo Clinic in Arizona (MCA). Four use cases illustrate how the RQD is used to assist the clinical practice. Also, a case study on how the DDD enabled an effective intervention for reducing lengthy radiologist turnaround times (TATs) for observation patients (ObP) is presented.

Topic 2: **Integration of Sparse Singular Vector Decomposition and Statistical Process Control for Monitoring and Change Detection of High-Dimensional Data Streams.** The establishment of DDD makes it possible to collect data on QoC metrics in real time. The data is high-dimensional if considering stratification of each QoC metric in terms of scanners, protocol types, patient types, and body parts. Collectively monitoring stratified QoC metrics instead of aggregated ones enables root cause identification of anomalies detected by the monitoring scheme. In another research project of mine, I developed a method for monitoring high-dimensional data streams that integrated Singular Vector Decomposition (SVD) and Multivariate Statistical Process Control (MSPC), called SSVD-MSPC. Monitoring of high-dimensional data streams has been a popular research topic in MSPC in recent years. However, existing methods have the limitation of mostly requiring a parametric distribution for the data, while non-parametric methods are lacking. Real-world data such as QoC metrics are rarely parametric. Also, the existing methods are not efficient enough to suit the need for real-time analytics of data streams that could be spatially high-dimensional and temporally high-throughput. The proposed SSVD-MSPC method addressed the aforementioned limitations of the existing methods and additionally provides three key capabilities, including monitoring, fault identification, and fault characterization. Extensive case studies are conducted for small, medium and large scaled data streams that experience faults of different magnitudes and various temporal shapes. SSVD-MSPC achieves universally good performance across the different settings in comparison with existing methods. To demonstrate the performance of proposed SSVD-MSPC algorithm, we modeled the radiologic care process as a discrete events systems for monitoring 51 QoC metrics in Radiology Department at Mayo Clinic, and designed three different types of common situations that would be considered as abnormal. The SSVD-MSPC successfully detected two out of three faults efficiently. For the third type of situation, the SSVD-MSPC helped us to find an optimal technician schedule plans based on our findings.

Topic 3: **Multi-Modality Deep Transfer Learning for Computer-Aided Diagnosis.** The aforementioned projects, Topic 1 and 2, focus on solving problems related to tracking HISs. The archive HIS is another type of widely deployed information systems. This topic focus on archive HISs in radiology department, PACS is a typical archive HIS which stores scanned medical images from multiple modalities such as computed tomography (CT), magnetic resonance imaging (MRI), X-ray and mammography. Different from the data in tracking HISs, information stored in archive HISs have unique characteristics: 1) the size of data is huge. Considering contrast-enhanced digital mammography (CEDM) as an example, each scan produces 4 high-resolution images with different scan parameters. Each of the images is around 30 MB which is about $1,000$ times larger than the data recorded in tracking HISs for the same procedure. If we consider all 4 images, the size difference is over $4,000$ times. 2) The raw data from an archive HIS are not straightforward to interpret, even by medical professionals. Lots of information is hidden underneath the visualization of images, and therefore techniques of computer vision and pattern recognition are need to help the interpretation. In this project, we collaborated with Mayo Clinic in

Arizona and focused on breast cancer image classification using CEDM images in PACS. Compared to conventional machine learning approaches, deep learning techniques showed promising performance on open-domain classification tasks. However, the uniqueness of our task makes it difficult to directly apply existing approaches: 1) since CEDM is a new imaging technique, the sample size used to train a classification model is not up the standard required by deep learning. 2) CEDM produces four images from two views and two energy frequencies for each patient. How to optimally integrate these images affects the classification performance. To address these challenges, we proposed a deep learning framework which adopts the ideas of transfer learning and multi-modality integration. The experimental results indicate good performance on CEDM data collected in Mayo Clinic. Also we provided recommendation on tumor segmentation in order to reduce the labor-intensive manual segmentation process.

1.3 Summary of Original Contributions

The objective of my dissertation research is to develop new methods that overcome the aforementioned limitations of the existing methods and demonstrate the utility of the methods in real applications with data and information collected from multiple types of HISs. In my methodological development, I focus on two major categories of data: radiologic process records from tracking HISs, and medical images from archive HISs.

The summary original contributions of my dissertation research are summarized as follows:

- Design of a super-HIS, DDD, by integration of multiple HISs, which allows for an end-to-end tracking of the care patients receive in radiology department from check-in to radiologic report generation. DDD provides the information infrastructure that makes it possible to accomplish multiple goals related to process improvement, such as quality of care monitoring, workflow analysis and productivity assessment, scheduling, equipment utilization and load balancing.

- Development of a non-parametric, computationally efficient method, SSVD-MSPC, for monitoring and change detection of high-dimensional high-throughput data stream. This method can be used in various application domains. My dissertation demonstrates the effective application of SSVD-MSPC in two applications: one is to monitor the quality of care metrics in radiology; the other is to monitor the quality of service in MCCN.
- Development of a multi-modality deep transfer learning framework for computer-aided diagnosis. The transfer learning technique allows for adopting knowledge gained from large existing dataset from other domains and fine-tuning model parameters using smaller but specific datasets in our focused domain. The multi-modality approach enables consideration of the intra-relations among all images from the same patient. My dissertation focuses on a demonstration of breast cancer classification using CEDM images.

The dissertation research will be presented in three chapters. Chapter 2 presents the development of topic 1: Integration of Multiple Health Information Systems for Quality Improvement of Radiologic Care. Chapter 3 presents the development of topic 2: Integration of Sparse Singular Vector Decomposition and Statistical Process Control for Monitoring and Change Detection of High-Dimensional Data Streams. Chapter 4 presents the development of topic 3: Multi-Modality Deep Transfer Learning for Computer-Aided Diagnosis.

Chapter 2

INTEGRATION OF MULTIPLE HEALTH INFORMATION SYSTEMS FOR QUALITY IMPROVEMENT OF RADIOLOGIC CARE

2.1 Introduction

Health care spending in the U.S. has been estimated to account for 17% of gross domestic product, nearly twice as much as that in other developed countries (Hartman *et al.*, 2015). In spite of this enormous expenditure, the U.S. ranked the last in health care quality among developed countries according to a 2008 Commonwealth Fund report (Roehr, 2008). Quality of Care (QoC), according to a 2001 report by the Institute of Medicine (Kohn *et al.*, 2001), includes six dimensions: timeliness, efficiency, effectiveness, patient safety, patient/family centeredness, and equity of care.

QoC improvement initiatives generally prioritize areas that incur the most expenditure. One such area is radiologic care as it involves the use and maintenance of expensive imaging equipment. There is ample evidence showing that Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) have contributed significantly to the rising cost of health care (Hu *et al.*, 2011). As a result, the Centers for Medicare and Medicaid Services have proposed that imaging devices costing greater than one million dollars should be amortized for replacement based on a 90% service utilization, in the hope of lowering per-patient reimbursement cost (Centers for Medicare & Medicaid Services , CMS).

To improve the QoC in radiology, an important first step is to define metrics to measure the QoC. Metrics are numerical indicators used to measure the performance in areas considered important for an organization's mission (Abujudeh *et al.*, 2010). Metrics for the QoC in radiology have been discussed in a number of papers. Typical examples include report turnaround time (Abujudeh *et al.*, 2010), patient access and wait times (Sarwar *et al.*, 2015), equipment utilization rates and downtimes (Sarwar *et al.*, 2015), exam volumes (Ondategui-Parra *et al.*, 2005), and staff workload (Ondategui-Parra *et al.*, 2005).

With the rapid development and adoption of information technology in health care, electronic Health Information Systems (HISs) have been widely used in health institutions. This has provided an unprecedented opportunity for acquiring quantitative data on the care process, from which metrics for QoC can be extracted efficiently and automatically. In radiology, several HISs are commonly used, including the Electronic Health Records (EHR), Radiology Information System (RIS), and Picture Archive and Communication System (PACS). Each of these HISs records partial and complementary information about the entire radiologic care process. Specifically, the EHR contains patient information and detailed medical history. Other than patient information that overlaps with the EHR, the RIS includes radiology-specific measures such as technologist imaging verification time. The PACS focuses on storing the digital images from the exams and related metadata information using the Digital Communications in Medicine (DICOM) standard (Association *et al.*, 1997). To the best of our knowledge, no research has been done to integrate the multiple HISs together to allow end-to-end tracking of the care each patient receives in radiology department, i.e., from check-in to finalization of the radiologic report. As a result, the current QoC research is limited as it is not capable of capturing the entire radiologic care workflow given only segmented data available from a single HIS. A comprehensive assessment of radiology QoC requires the multiple HISs be integrated such that various key QoC metrics can be extracted from the resulting “super-HIS”.

At Mayo Clinic, we have a home grown radiology quality assurance system called Dose Index Tracker (DIT©) (Wang *et al.*, 2011). The DIT system collects information directly from the DICOM headers of scanned images. Such information includes but is not limited to scanner information (ID, vendor, etc.), exam related information (procedure, timestamps of each scanned image, radiation dose, etc.), and basic patient information. The DIT was designed to track patient-specific radiation dose across all radiology exams performed at Mayo Clinic, and to provide intelligent data analysis, reporting, and alerting in a clinical quality assurance context.

We present our development of a technology called Department Data Depot (DDD) that integrates four HISs, including an EHR, a RIS, a PACS, and the home grown DIT system. In the development of DDD, we adopt the concept of loose-coupling techniques in database integration and propose a three-layer integration framework, including a data mashup lower-layer, an aggregation service middle-layer, and a result presentation upper-layer. The loose-coupling architecture is a mature and well-known technique which is designed to reduce the risk that a change made within one or more database will create unanticipated changes within other related databases. Limiting interconnections can help isolate problems when things go wrong and simplify testing, maintenance and troubleshooting procedures. As a tradeoff, such integration may slightly increase the response time of the system and necessitate extra maintenance of the mid-layer due to the nature of design. In the data mashup layer, a module is introduced to maintain the relations and constraints among the integrated HIS databases. When any of the source databases or the relations and constraints themselves change, we can easily alter the mid-layer of our integration to leave the upper levels untouched.

In addition to the design and implementation of this framework, a significant amount of effort has been spent on addressing specific issues from incompatibility of the multiple HISs, such as inconsistent data fields, data measurement errors, missing values, and human errors. These issues could substantially affect the usability of the integrated system and therefore have been deliberately addressed in our research. Furthermore, we propose nine QoC metrics defined upon the integrated system: (1) exam duration, (2) technologist post-processing time, (3) technologist turnaournd time (TAT), (4) radiologist TAT, and (5) total TAT, which reflect the timeliness and efficiency of radiologic care; (6) patient waiting time and (7) patient TAT, which reflect the efficiency and patient satisfaction; (8) patient volume and (9) exam volume, which reflect the workload distribution. All these metrics measure QoC from different and complementary perspectives. Finally, we present the deployment of DDD in radiology department of Mayo Clinic in Arizona (MCA) through two case studies. DDD is deployed in MCA through a web portal, called Radiology Quality Dashboard (RQD). In the first case study, we demonstrate, through four examples, how users can use RQD information in the clinical practice. In the second case study, we show how DDD enabled identification of the root cause of lengthy radiologist TAT for observation patient (ObP) – a specific patient subtype, and further enabled the development of an effective intervention for radiologic quality improvement.

Different from data integration and quality control applications in other fields, such systems in healthcare have restricted access policies to protect patient information according to HIPPA (Health Insurance Portability and Accountability Act). Indeed, our research team had to go through HIPPA training before we implemented the project, and the end users of the project are from Mayo Clinic in Arizona and have the right to access the information. When our DDD system acquires raw data records for other HISs, those outer source systems give us the privilege to query and store patient information. When we use our database to calculate quality metrics and generate reports, however, an anonymization procedure is applied in the mash up layer by removing all patient related information except the patient type, such as “inpatient”, “outpatient” and “ED patient”. Also, the metrics of interest in this research are aggregated measures; they are not specific to an individual patient. As a result, the patients’ demographics and disease information are not used in the analysis.

The contributions of this paper are multifold:

- Our work is the first of its kind and provides a technology for multi-HIS integration for radiology practice. By integrating these HISs together, DDD enables end-to-end tracking of the radiologic care each patient receives, with detailed time stamps and contents of each care activity as well as rich information on patients, providers, and equipment. While this paper focuses on quality improvement, the data and information in DDD can support a variety of other goals including, but not limited to, scheduling, load balancing, and process optimization. In this sense, we envision that DDD has a potential for profoundly impacting radiology practice.
- Based on intensive interaction and dialogue with radiologists, technologists, and administrators in radiology department cross-referenced with the available data in DDD, we propose nine QoC metrics that are important for monitoring, tracking, and evaluating the quality of radiologic care. These metrics have not previously been available.
- DDD was deployed in MCA in September 2015. Since then, it has been used extensively by clinicians, administrators, and researchers to monitor QoC, identify problem areas, and perform interventions to improve the quality of radiologic

care. In this sense, the research in this paper sets an example that close collaboration between industrial engineers and clinicians has great potential for transforming health care practices.

2.2 Literature Review

The recent, widespread adoption of HISs in health institutions has made it possible to collect detailed, quantitative data on the care process. Availability of the data further enables QoC to be measured and improved. In this section, we will review major types of HISs, focusing on how they have been used in relation to QoC, especially in radiology.

The terms of EHR or Electronic Medical Records (EMR) are often used interchangeably with the HIS. In some care settings, the EHR is the only HIS in play. In this paper, we use EHR to refer to the enterprise-level HIS of patient medical history. The EHR includes all key administrative clinical data relevant to a patient's care pulled from multiple encounters and facilities, such as demographics, progress notes, medications, immunizations, vital signs, laboratory and radiologic reports. The EHR was one of the earliest HISs and was in use when health care practices began transitioning from the paper era to the digital era. Early research on the EHR in relation to quality focused on design and implementation issues of the EHR to make it a proper information enabler. For example, (Walker *et al.*, 2008) proposed a coordinated set of steps for safe design, implementation and improvement of the EHR. (Middleton *et al.*, 2013) made ten recommendations on the EHR with respect to improving the safety and quality of care. (Wang *et al.*, 2003) performed a cost-benefit analysis of the EHR in ambulatory primary care settings and concluded that the EHR investment had a positive financial return. (Miller *et al.*, 2005) conducted similar case studies on 14 solo or small-group primary care facilities and suggested that the EHR would be "financially attractive" for some facilities and "financially acceptable" for most others. More recently, as the EHR was becoming a mature and widely adopted system, research has focused on how the data and information recorded by the EHR can be used for quality and performance improvement of health care (Amoah *et al.*, 2015). For example, (Poissant *et al.*, 2005) studied the documentation time of physicians and nurses, and concluded that with the help of the EHR, nurses save about 24% of the overall time spent on documenting during a shift. (McVeigh *et al.*, 2008) proposed several metrics to be extracted from the EHR that measure the timeliness, an important dimension of QoC, of several sub-processes of optometry practices, such as check-in, pretesting, doctor examination, and optical sub-processes. Despite the wealth of existing studies, little research has been conducted on quality improvement of radiologic care using the EHR. This is because the EHR is an enterprise-level system such that the data it collects lacks sufficient granularity to help extract radiology-specific QoC metrics such as pre-radiologic-exam patient waiting time, exam duration, and radiologic report turnaround time.

The RIS and PACS are two specialized HISs for radiology. Radiology departments use a RIS to track patients, exams, result distribution, and procedure billing. A PACS provides economical storage and convenient access to images from multiple modalities. The PACS stores both image data as well as image related data-flow information. In this section, only information from data-flow are discussed. (Wang *et al.*, 2011) developed a DIT©to extract, store, and monitor critical radiation dose indicators stored in DICOM file headers found in PACS. (Hu *et al.*, 2011) developed five metrics for efficiency benchmarking, including exam duration, inter-series time, inter-patient time, appointment interval time, and table utilization using DICOM information stored in PACS. A radiology department's RIS and PACS are generally designed to interface and can often be easily linked to provide more comprehensive information than using a single system alone. Research has been done to use RIS and PACS together for improving the quality and safety of radiologic care. For example, (Nitrosi *et al.*, 2013) developed a procedure to use Health Level 7 (HL7) standard messaging in RIS and PACS to reduce clinical risks due to patient reconciliation errors. In several independent studies, researchers developed various tracking systems with data from RIS and PACS to monitor the overall performance and exam status within radiology department in order to improve patient satisfaction and outcome assessment (Nagy *et al.*, 2009). (Seltzer *et al.*, 2000) integrated RIS, hospital information systems, and manually input data to extract several management metrics such as report turnaround time, access to appointments, and productivity.

As seen, most existing research was based on a single or department-level HIS. However, radiologic care is a complex process such that data describing the entire care process resides in multiple HISs. For example, patient check-in time, type, and demographics are available in EHR. Service time stamps such as the times when imaging was started and finished, and when the image was verified by the technologist are stored in RIS. The time when the radiologic report is finalized by the radiologist is recorded in PACS. Image files together with meta data such as modality, body part, and with/without contrast are also stored in PACS. As a result, although a few QoC metrics may be extracted from a single HIS alone, these metrics only provide partial, limited information about the QoC. A comprehensive assessment of the QoC in radiology requires the multiple HISs be integrated into a "super-HIS" from which various key QoC metrics can be extracted. Without the integration, many important

QoC metrics that require linked records from multiple HISs would be missed. For example, an important QoC metric is patient pre-exam waiting time, i.e., the time duration between check-in and imaging start. The two time stamps needed to compute the waiting time reside in EHR and RIS, respectively. Another important QoC metric, patient turnaround time, is measured by the difference between two time stamps, i.e., imaging start and radiologist completion of reading the image and finalizing the report, which are in RIS and PACS, respectively. Furthermore, to measure the distribution of the aforementioned time metrics as well as other QoC metrics such as patient volume and exam volume with respect to different patient types, imaging modalities, facilities/sites, scanners, and body parts/sub-specialties, information needs to be pulled from EHR or PACS to group-partition these metrics.

2.3 Development of Multi-HIS DDD and Radiologic QoC Metrics

In this section, we present our development of DDD that integrates multiple HISs. We also define and describe how we extract a collection of key QoC metrics from the DDD.We will present our research development in the context of radiology department at MCA, but the developed technologies are generalizable to other health institutions.

2.3.1 Mapping out Radiologic Care Process and Interrogation of the Multi-HIS

Before developing the DDD and extracting the key QoC metrics, we needed to identify the major steps involved in the care process performed within radiology department for patients. Through observations and intensive dialogue with the radiologists, technologists, and administrators in radiology department of MCA, we

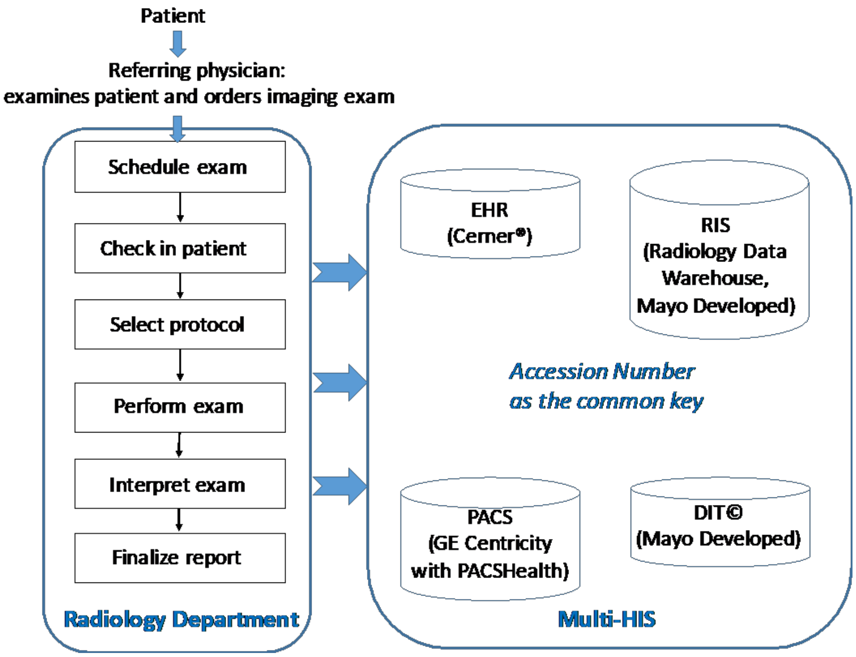


Figure 1: Radiologic Care Process and Multiple Relevant HISs

mapped out the radiologic care process as depicted in the left hand side of Figure 1. Furthermore, we dove into each HIS used in radiology to identify what information about the mapped care process was stored in the HIS. There are three important observations: (1) No single HIS provides end-to-end measurement for the entire care process. (2) Each of the four HISs in the right hand side of Figure 1 contains useful while unique information required to describe the entire care process (please see Figure 4 for details). This suggests that all four HISs must be included in developing the DDD. (3) There exists a common data field across all four HISs, i.e., the accession number, which is a unique identifier for each exam of each patient (one patient can have multiple exams). The accession number can be used as a key to link the four HISs together to track the entire care process on a per-patient per-exam basis. These findings lay the ground work for the development of DDD.

2.3.2 DDD Architecture

DDD integrates four major HISs deployed at MCA: an EHR – Cerner ®, a RIS – Radiology Data Warehouse (RDW), a PACS monitoring system – PACSHealth that extracts exam status changes in GE Centricity PACS, and a custom-built radiation dose tracking system – DIT ©. Database integration is an important technique that helps the data users interrogate heterogeneous records, information and relationships among multiple data sources and provides a unified data view. One traditional techniques are called data warehousing (Inmon (2005)), which extracts data from multiple sources, transforms the data into a proper and unified format, and then loads the data into another standalone target for further query and analysis. A major limitation of this technique is the tight relationship to the original data sources, which makes it difficult to adopt any upstream structural changes and increases maintenance/update costs (Moseley (2009), Wu *et al.* (2007)). More recently, loose-coupling techniques have been proposed for database integration, which provide a unified real-time data query interface over a target data source (Kaye (2003)). Such techniques are developed and used as an important part of Service Oriented Architecture (SOA) (Erl (2008)). Loose-coupling techniques rely on mappings between the data structures of the original data sources and the target data source. If required, transformation techniques are used to wrap the interfaces of original sources for a higher-level query. Depending on the mapping schemas, the techniques can be categorized into two basic types: Global As View (GAV), which maps records in the target data source to original data sources; Local As View (LAV), which maps records in the original data sources to the target data source. Our development of DDD adopts the concepts of loose-coupling techniques by creating higher-level data schema with LAV mapping. The original data

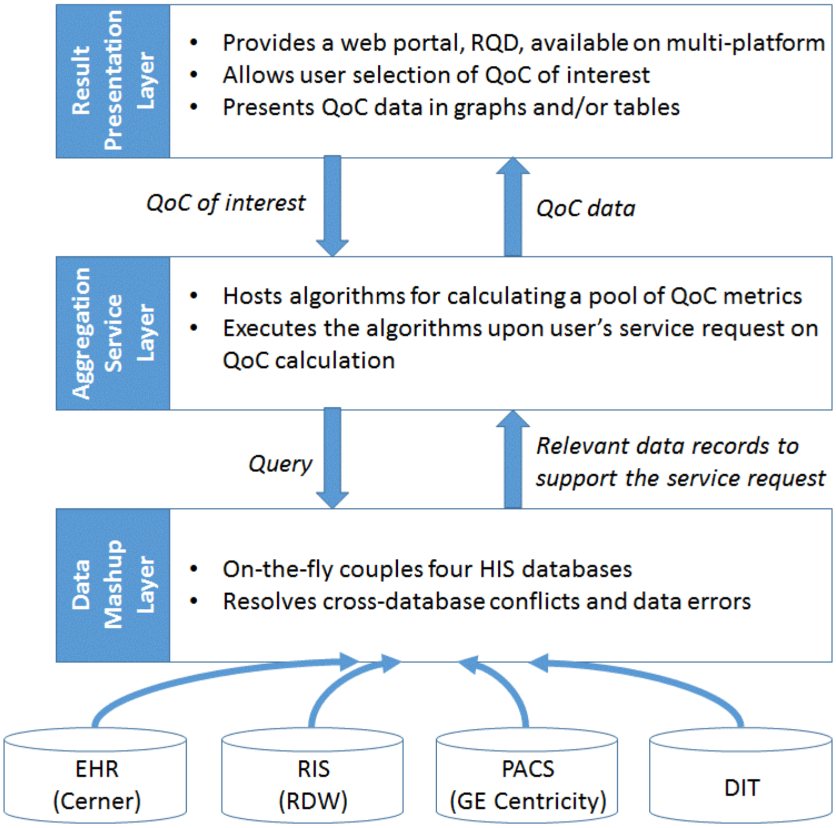


Figure 2: Three-Layer Integration Framework of DDD

records are unchanged and connected dynamically to construct a local record for data analysis. A significant challenge we encountered was how to resolve semantic conflicts among the different data sources, as heterogeneous definitions and/or meanings always exist when multiple data sources are to be linked together. To tackle this challenge, semantic and ontology-based integrations are developed by involving expert knowledge that explicitly defines schema terms.

Specifically, we propose a three-layer integration framework for DDD, including a data mashup layer, an aggregation service layer, and a result presentation layer, as shown in Figure 2. Next, we introduce each layer with more details.

2.3.2.1 Data Mashup Layer

The data mashup layer couples the data fields in each individual HIS into the DDD, on the fly. The coupling uses “accession number” as a unique identifier for each exam of each patient and a common field shared by all the individual HISs. Each record in DDD corresponds to one exam. Each exam is associated with a collection of attributes to describe it, which correspond to the joint data fields from the four individual HISs. Here, only relevant ones associated with QoC metrics are coupled in DDD (see Table 1).

In review of Table 1, we note that these HIS databases have heterogeneous data fields, which may be inconsistent and even conflict with each other. Taking facility or site as an example: the “Facility” data field from RDW is an indicator for the two sites of MCA (hospital vs. clinic); the “STATION_ID” from DIT can also help identify the site of each exam, as it is the unique ID of each scanner. Although representing the same concept, the data fields are named differently across different HISs. A second issue is that several data fields are input by staff manually, which may introduce human errors. To deal with such inconsistencies and errors in data mashup, we developed three heuristic rules based on intensive dialogues with the radiology staff.

Heuristic Rule I: technologist finish time t_{tech_finish}:

t_{tech_finish} is the timestamp when a technologist finishes the exam procedure on a patient. At this time, all scanned images are ready to be sent to the radiologist to

Table 1: Relevant Data Fields in Each HIS to QoC Metric and Included in DDD

EHR (Cerner)

Field Name	Description
MRN	Medical record number unique to patient.
ACCESSION_NBR	Primary key to link individual HIS, e.g., “810713204-1”.
CHECKIN_DT_TM	Timestamp when the patient checks in, e.g., “2016-05-24 10:34:33.000”.
ENCNTR_TYPE_DISPLAY	Encounter type – an indicator of patient type, e.g., outpatient, inpatient.
RESOURCE_BEG_DT_TM	Timestamp when the exam is scheduled, e.g., “2016-08-01 10:15:00.000”.
STATE_MEANING	Current status of the exam, e.g., “CHECKED IN”, “CONFIRMED”.

RIS (RDW)

Field Name	Description
AccessionNumber	Primary key to link individual HIS, e.g., “810713204-1”.
Code	Procedure code, e.g., “70030K”, “74364”.
Encntr	Encounter type – an indicator of patient type, e.g., outpatient, inpatient.
Facility	Facility / site where the exam is taken, e.g., “hospital”, “clinic”.
Modl	Imaging modality, e.g., “MRI”, “CT”.
ServiceTime	Timestamp when the image scanning finishes, e.g., “2016-07-18 10:12:32.000”.

PACS (GE-Centricity)

Field Name	Description
AccessionNumber	Primary key to link individual HIS, e.g., “810713204-1”.
FiftyTransition	Timestamp when the technologist finishes verifying the exam, e.g., “2016-07-18 10:12:32.000”.
MedicalRecordNumber	Medical record number – a patient ID, e.g., “12345678”
Modality	Imaging modality, e.g., “MRI”, “CT”.
NinetyTransition	Timestamp when the radiologist finishes dictating the exam, e.g., “2016-07-18 10:12:32.000”.
ProcedureCode	Procedure code, e.g., “70030K”, “74364”.
TwentyTransition	Timestamp when the exam is ordered, e.g., “2016-07-18 10:12:32.000”.

DIT (Mayo)

Field Name	Description
AccessionNumber	Primary key to link individual HIS, e.g., "810713204-1".
FINISH_TIME	Timestamp when the last image of an exam arrives at PACS, e.g., "2016-07-04 18:23:00.000".
PROTOCOL_NAME	Exam protocol name, e.g., "8a_RENAL_DONOR".
START_TIME	Timestamp when the first image of an exam arrives at PACS, e.g., "2016-07-18 10:12:32.000".
STATION_ID	Unique scanner ID, e.g., "JA_CT0PRA81818".

dictate and the patient will be transferred to the recovery room. t_{tech_finish} is needed for deriving an important QoC metric, technologist TAT. t_{tech_finish} does not exist in any of the four HISs, but is indirectly measured by two data fields: "ServiceTime" in RDW and "FiftyTransition" in PACS. The former is the time when the technologist manually indicates that scanning has finished. In contrast, the "FiftyTransition" marks the time when the images are marked as "Verified" in PACS indicating that all image processing is complete. Typically, "ServiceTime" is earlier than "FiftyTransition" because it does not include the time the technologist spends on post-processing the scanned images. However, since the "ServiceTime" is manually entered into the system, depending on each technologist's working habit, this timestamp may be earlier or later than the exact t_{tech_finish}. To eliminate the bias of the input and obtain a more accurate t_{tech_finish}, we use the following rule:

$$t_{tech\_finish} = later(\text{“FiftyTransition”}, \text{“ServiceTime”}) \tag{2.1}$$

The rationale behind this is that if "FiftyTransition" is later than "ServiceTime", it means that the technologist inputs the "ServiceTime" right after the scanning is finished but does not consider the post-processing time. Therefore, in such situation, "FiftyTransition" is a better measure for t_{tech_finish}. On the other hand, if "FiftyTransition" is earlier than "ServiceTime", there are likely extenuating circumstances or additional patient interactions that increase the hands-on component of the exam. In this case, "ServiceTime" is a more appropriate time stamp for t_{tech_finish}.

Heuristic Rule II: check-in time t_{check_in}:

for patients in Emergency Department (ED): t_{check_in} is needed for deriving an important QoC metric, patient waiting time. It is measured by "CHECKIN_DT_TM" in Cerner. However, the "CHECKIN_DT_TM" is missing for patients in ED due to the unique care process of ED. A patient's radiology "check-in" (or alert of arrival) from the ED happens when an imaging exam is ordered by the physician, which is stored in the field of "TwentyTransition" in PACS. Therefore, we use "TwentyTransition" as t_{check_in} for ED patients, which produces the following rule:

$$t_{check\_in} = \begin{cases} \text{“TwentyTransition”} & \text{if the patient is in ED} \\ \text{“CHECKIN\_DT\_TM”} & \text{otherwise} \end{cases} \tag{2.2}$$

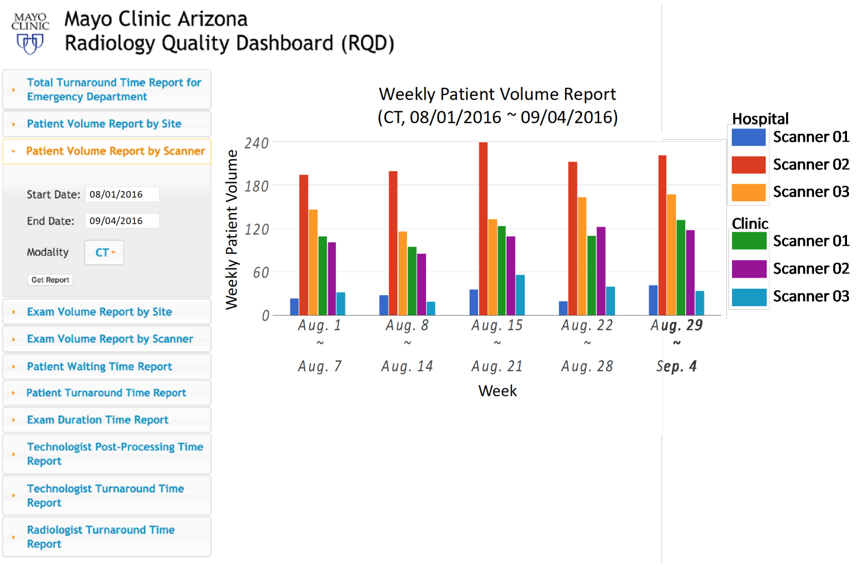
Heuristic Rule III: patient type classification:

It is important to be able to compute a QoC metric for different patient types such as inpatient, outpatient, and ED patients. This would help reveal QoC problems in serving each type of patient and properly allocate resources to overcome the problems. Patient type is stored in "ENCNTR_TYPE_DISPLAY" in Cerner and "Encntr" in PACS. Unfortunately, it is observed the two data fields have a large number of missing values (i.e., NULL values). To mitigate the problem, we use "ENCNTR_TYPE_DISPLAY" as the primary source to obtain the patient type, because its missing data problem is less severe than “Encntr”. When the “ENCNTR_TYPE_DISPLAY” is missing for a patient, we check “Encntr”. If “Encntr” is also missing, we label the patient as “NA”.

$$
\text{patient type} =
\begin{cases}
\text{ED patient} & \text{if ``ENCNTR\_TYPE\_DISPLAY'' = `Emergency' OR} \\
 & \text{``ENCNTR\_TYPE\_DISPLAY'' = NULL but ``Encntr''=`EM'} \\
\text{inpatient} & \text{if ``ENCNTR\_TYPE\_DISPLAY'' = `Inpatient', `Observation' OR} \\
 & \text{``ENCNTR\_TYPE\_DISPLAY'' = NULL but ``Encntr''=`IP'} \\
\text{outpatient} & \text{if ``ENCNTR\_TYPE\_DISPLAY'' = `MCA Hospital C',} \\
 & \text{`MCA Patient', `OP in a bed',`Pre-Admit Outpatient',} \\
 & \text{`Recurring AIC', `Recurring PM\&R', `Recurring Rad Onc' OR} \\
 & \text{``ENCNTR\_TYPE\_DISPLAY'' = NULL but ``Encntr''=`OP', `P'} \\
\text{NA} & \text{otherwise}
\end{cases}
\tag{2.3}
$$

2.3.2.2 Aggregation Service Layer

This layer hosts the algorithms to derive the nine QoC metrics (please see Section 3.3 for details). Execution of an algorithm is triggered by the user’s service request on the corresponding QoC metric, together with a time interval and strata that the user wants to utilize in the computation of the QoC. The QoC metric can be stratified by patient type, facility site, and scanner. Once an algorithm is triggered, it will query DDD, perform filtering and arithmetic operations, and return the user-requested QoC measurement that is presented on the result presentation layer (see below).



Start Date: 08/01/2016

End Date: 09/04/2016

Modality CT

Figure 3: Result representation layer – a web portal, RQD

2.3.2.3 Result Presentation Layer

We built RQD, a web portal as the result presentation layer. RQD adopts HTML 5 techniques and provides users easy access to the aggregation service layer from desktop computers, laptops, smart phones, and tablets. A snapshot of RQD is shown in Figure 3. In particular, on the left side of RQD, a user can select the QoC metric of interest, start and end dates and strata for which the QoC is to be computed. This information is sent to the aggregation service layer and results are presented on the right side of the RQD as graphs and/or tables.

2.3.3 Definition of Radiologic QoC Metrics

Based on the radiologic care process mapped out in Section 2.3.1, we define nine QoC metrics. Specifically, we propose five metrics measuring the timeliness

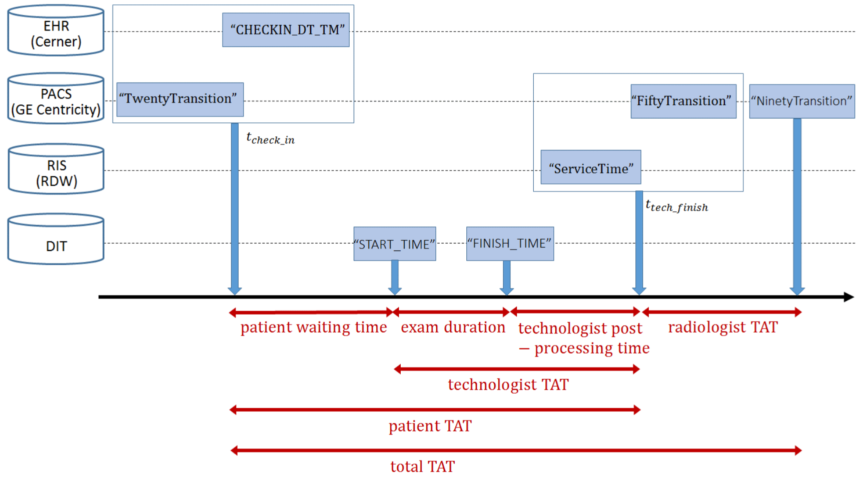


Figure 4: Timestamps used in the formula for calculating the QoC metrics and the HISs each timestamp comes from

and efficiency of radiologic care: exam duration, technologist post-processing time, technologist TAT, radiologist TAT, and total TAT. We propose two metrics on efficiency and patient satisfaction: patient waiting time and patient TAT. In addition, we propose two metrics on measuring the demands and workload: patient volume and exam volume. Table 2 provides the definition and formula of each QoC metric. Figure 4 further shows the relative positions of the timestamps used in the formula (last column of Table 2) and from which HISs each timestamp can be obtained.

2.4 Application of DDD and QoC Metrics in Quality Improvement of Radiologic Care

In this section, we present the applications of DDD technology in radiology department at MCA. The first application demonstrates how RQD enabled by DDD was used to retrieve important information in order to help identify areas of improvement for radiologic care quality. The second application demonstrates how DDD enabled

Table 2: Definitions and formula of the proposed QoC metrics

QoC Metrics	Definition	Formula
exam duration	Time duration of the scanning process	“FINISH_TIME” – “START_TIME” in DIT
technologist TAT	Time duration for a technologist to perform the exam	t_{tech_finish} in (2.1) – “START_TIME” in DIT
technologist post-processing time	Time duration for a technologist to verify the images, write exam notes, and clean up the exam room after the scanning process ends. This is a portion of technologist TAT.	t_{tech_finish} in (2.1) – “FINISH_TIME” in DIT
radiologist TAT	Time duration for a radiologist to read the images and complete the diagnostic report, after receiving the images from the technologist	“NinetyTransition” in PACS $-t_{tech\_finish}$ in (2.1)
total TAT	Time duration between the patient check-in time and report completion by the radiologist. This is the total time spent for the radiologic care of a patient.	“NinetyTransition” in PACS $-t_{check\_in}$ in (2.2)
patient TAT	Time duration between the patient check-in and check-out times. This is the total time for a patient’s physical stay in radiology department	t_{tech_finish} in (2.1) $-t_{check\_in}$ in (2.2)
patient waiting time	Time duration between patient check-in time and the start of the scanning process. This is the inactive portion of a patient’s physical stay in radiology department.	“START_TIME” in DIT $-t_{check\_in}$ in (2.2)
patient volume	Number of patients	Unique Count(MRN)
exam volume	Number of exams (one patient may have multiple exams)	Unique Count(AccessionNumber)

the identification of the root cause of lengthy radiologist TAT for a specific patient subtype, observation patients (ObP), and further enabled the development of an effective intervention for radiologic quality improvement.

2.4.1 RQD and its clinical use cases

With the DDD technology and nine QoC metrics, a number of radiologic care quality related questions can be answered. Here, we present four examples on how RQD could potentially help improve care. At the time of preparing this paper, we chose to select five full weeks' data (August 1, 2016 – September 4, 2016) and used CT as the example for illustration purposes. The same applies to other modalities such as MR.

2.4.1.1 Example I

This study was motivated by a concern raised by the ED that the turnaround time of CT after regular radiologist working hours was longer than expected. As shown in Table 2, the turnaround time is defined as the duration between check-in time and radiologist finish time. CT is an imaging modality extensively used by the ED. CT exams are typically interpreted by attending radiologists during regularly working hours (radiology hours) and by residents and fellows during extended working hours (non-radiology hours). In response to a request from the ED, we used DDD together with RQD to investigate this perceived problem.

As shown in Figure 5, several observations can be obtained. First, the weekly average total TAT for both radiology and non-radiology hours ranges from 01:15 –

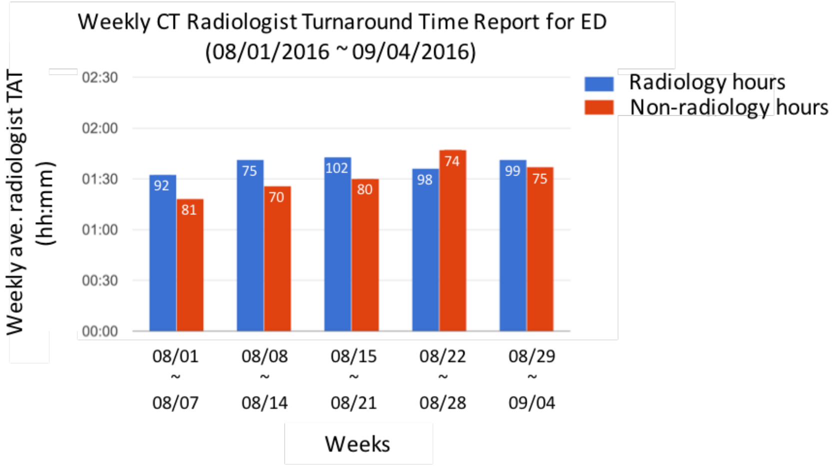


Figure 5: Example I - Weekly average total TAT during radiology and non-radiology hours (number in the bar indicates the exam volume)

01:45 (hours:minutes), which is reasonable (Wang *et al.*, 2015). Second, the total TAT during radiology hours is in general slightly longer than non-radiology hours suggesting the perception of afterhours delays was unwarranted. Given the data, we performed hypothesis testing to see if the observed difference between the radiology and non-radiology hours is statistically significant. The p-values for the five weeks shown in Figure 5 are 0.0135, 0.01229, 0.04456, 0.9193, and 0.2905, respectively. For the first three weeks in the selected range, the hypothesis tests support our observation that the total TAT during radiology hours is greater than that during non-radiology hours. However, the p-values of last two weeks are not significant. This may be interpreted as a result of residents and fellows joining the medical program in late July. Staff radiologists may provide greater assistance during the initial startup weeks as trainees become familiar with the radiology practice. And the increment of TAT should have other reasons, for instance, less radiologists in the last two weeks. Through this investigation, we concluded that the ED's concern regarding excessive TAT during non-radiology hours may not be valid overall. Yet, radiology department may need to

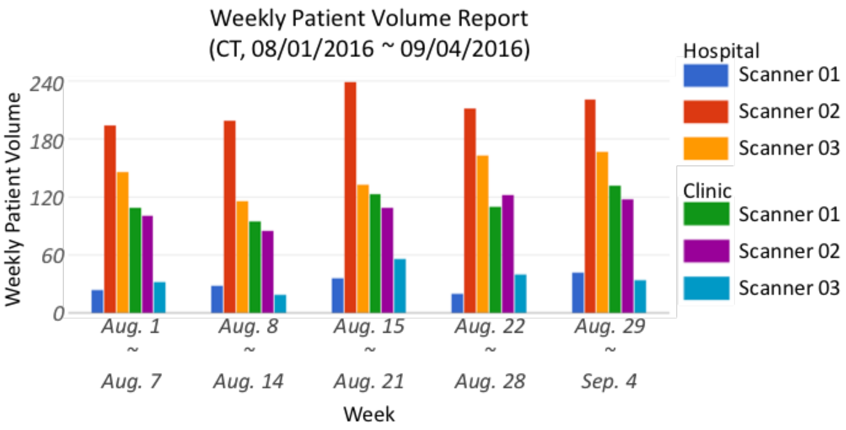


Figure 6: Example II - Weekly patient volume by scanner

pay special attention to the period when new residents and fellows are first taking responsibility for overnight call.

2.4.1.2 Example II

This study was motivated by the need for assessing the workload distributions among different scanners in order to better allocate resources and optimize scheduling. There are six CT scanners in MCA, with three located in the hospital and the others located in the clinic. The patient volume and exam volume (see Table 2 for details) are both reasonable indictors of scanner load.

Again, several observations can be obtained from Figure 6. First, it is clear that the three scanners in the hospital have unbalanced loads with the second scanner being heavily used while the first scanner being used substantially less. An ANOVA test was conducted to check if the patient volumes from three hospital scanners are all the same. The test results confirmed that there is statistical significance in the load imbalance across the three scanners at the hospital (p-value < 0.001). Second, the three scanners in the clinic also have unbalanced loads to some extent, although the issue is not as severe as the hospital. To confirm this, another ANOVA test was performed, which yielded p-value < 0.001, indicating that the three scanners in the clinic also have statistically significant load imbalances. Third, the overall load of scanners in the hospital is heavier than that in the clinic. A one-side two-sample t-test was conducted with the null hypothesis $H_0 : \mu_{hospital} = \mu_{clinic}$ and the alternative hypothesis $H_1 : \mu_{hospital} > \mu_{clinic}$, where $\mu_{hospital}$ and μ_{clinic} denote the average weekly patient volume in the hospital and the clinic respectively. The t-test yielded p-value $= 0.03526$, which indicated that we rejected the null hypothesis and the patient volume at the hospital is statistically higher. This is likely due to longer hours of operation in the hospital versus the clinic. However, some of the other imbalances such as the difference between Hospital Scanner 02 and 03 are more difficult to explain. It is our intention to explore this further with our clinical partners.

2.4.1.3 Example III

In patient care, patient waiting time reflects process efficiency and is also an important factor that affects patient satisfaction. This study is to assess the patient waiting time related to the radiology exams. To measure the patient waiting time, as defined in Table 2, requires the algorithm to know the patient check-in time (heuristically derived from EHR and PACS records with rule II, as shown in Figure 4) and exam start time (DIT records, shown in Figure 4).

Figure 7 shows the histogram of patient waiting time. The average waiting time is about one hour; 90% of the exams have patient waiting time less than two hours; five patients waited for more than three hours (the reasons for these extreme cases are

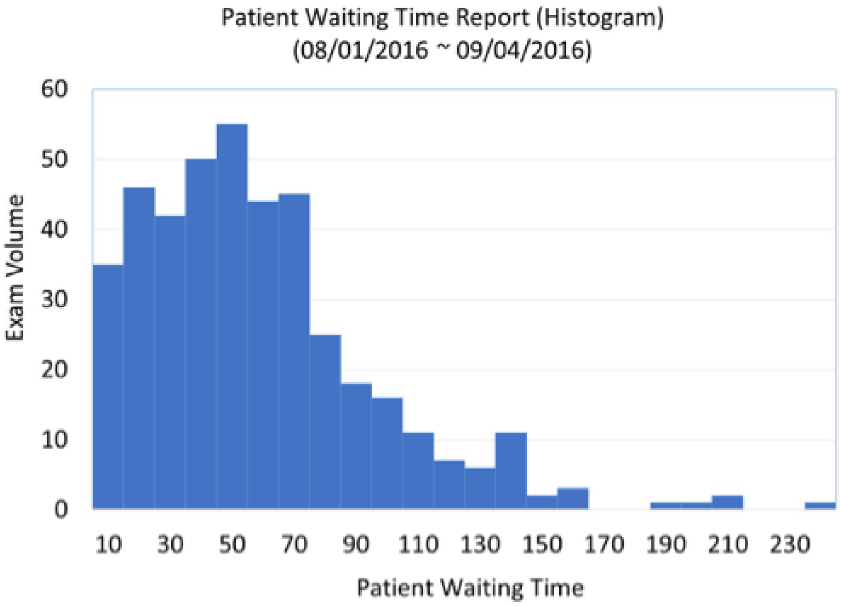


Figure 7: Example III - Histogram of patient waiting time

yet to be explored). Patient waiting time is a complex issue. It was observed that often times, patients check in earlier than their scheduled times, which leads to long waiting time. Also, waiting time is related to nursing assessment and/or oral contrast (for some CT exams) administration. This is an area that deserves more attention from radiology administration and more in-depth explorations.

2.4.1.4 Example IV

The last two steps in radiologic care are related to the activities from technologists and radiologists –two major service providers. Their TATs are important quality indicators. This study is to assess the technologist TAT and radiologist TAT. Both metrics share a timestamp $-t_{tech\_finish}$ and multiple data sources are involved (as shown in Figure 4 and Table 2).

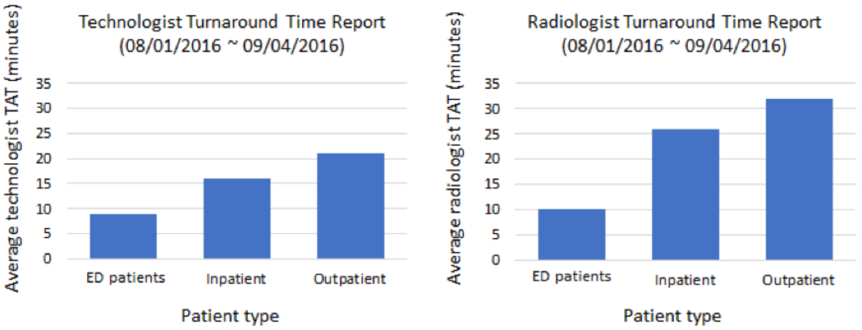


Figure 8: Example IV - Average technologist TAT (left) and average radiologist TAT (right)

Figure 8 shows a common trend shared by the technologists and radiologists, i.e., their TATs for outpatients are the longest, followed by inpatients and then ED patients. This trend is consistent with the urgency and typical complexity of care for these three types of patients. The average technologist TAT for outpatients, inpatients, and ED patients are 21, 16, and 9 minutes, respectively. The average radiologist TAT for outpatients, inpatients, and ED patients are 32, 26, and 10 minutes, respectively. While these numbers fall into a reasonable range, clinicians and administrators may still seek for ways to further reduce the TAT and improve the quality of radiologic care.

2.4.2 DDD-enabled intervention for improving radiologist TAT of observation patients (ObP)

ObP is a subtype of patients who have a condition for which the cause of symptoms is not immediately clear, so they are kept in the hospital for 23 hours to be monitored and/or to run more tests. From a workflow perspective, ObP is considered an inpatient because the patient's exam needs to be interpreted with priority. However, from a billing perspective, ObP is considered an outpatient.

In the fall of 2015 (before DDD was in place), radiology department received complaints from ordering physicians that the exams of ObP were not being interpreted in a timely fashion, i.e., these exams tended to have overly long radiologist TAT that did not match with the urgency level of ObP. Radiology department conducted an investigation but the root cause of this problem was not clear. In the second quarter of 2016 after DDD was deployed, the investigation was resumed. The root cause of the lengthy radiologist TAT for ObP was found to be that, prior to DDD, patient type classification was based on a single HIS, RDW, in which ObP were classified as outpatients. As a result, the exams of ObP did not appear on the radiologists' worklist for priority review. This caused delays in interpreting ObP exams by the radiologists. By integrating multiple HISs, DDD enabled patient type classification with more granularity, which led to ObP being separated out from outpatients as a standalone patient subtype. Leveraging this capability provided by DDD, radiology department started an intervention in the second quarter of 2016. A computer program was modified to automatically identify ObP exams from DDD and push those exams to the front of radiologists' worklist to be interpreted with priority.

To measure the effectiveness of the intervention, we collected data before and after the intervention. We focused on digital X-rays (Computed Radiology (CR) exams) of ObP interpreted by residents on Saturdays, since ordering physicians had previously complained about the lengthy TAT for these exams. Radiology department also wanted to exclude the possibility that the problem was related to the residents themselves. We queried DDD and obtained data from the first quarter of 2016, i.e., before the intervention took place. This included 101 CR ObP exams interpreted by residents on

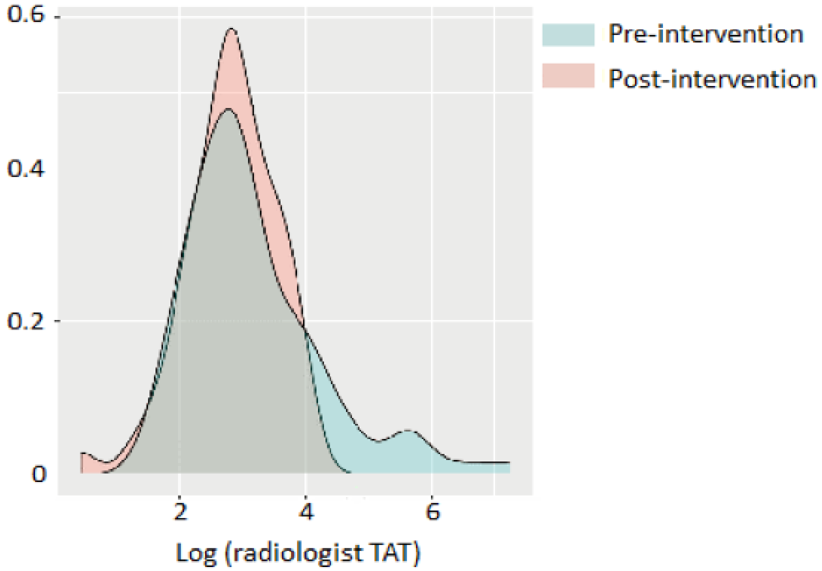


Figure 9: Probability distributions of logarithm of radiologist TAT for CR exams of ObP interpreted by residents on Saturdays

Saturdays. We also queried DDD to obtain data between 06/16-08/10/2016, i.e., after the intervention, which included 58 exams. We computed radiologist TAT as defined in Table 2 on each exam. Figure 9 shows the probability density plots on radiologist TAT for pre- and post-intervention exams. It is observed that the pre-intervention distribution of radiologist TAT has a heavy right tail, indicating that there is a non-negligible portion of exams with lengthy TAT. This problem is not seen in the post-intervention distribution. Furthermore, to verify the statistical significance of this finding, we performed a non-parametric two-sample proportion test for the following hypotheses:

$$H_0 : p_{pre}(T) = p_{post}(T) H_1 : p_{pre}(T) > p_{post}(T)$$

where $p_{pre}(T)$ and $p_{post}(T)$ denote the proportions of exams with radiologist TAT greater than T minutes pre- and post-intervention. The p-values for $T = 30, 60, 90$ minutes are 0.512, 0.005, and 0.033, respectively. This result implies that the intervention significantly eliminated ObP exams that took the residents longer than one hour to complete. Furthermore, on average, the intervention cut radiologist TAT by more than half, i.e., from 59.40 minutes pre-intervention to 20.67 minutes post-intervention. These results demonstrate that the intervention accomplished its intended purpose by reducing the radiologist TAT of ObP. Also, we eliminated the concern that the problem was caused by residents' work quality.

2.5 Discussion and Conclusion

In this project, we developed a novel technology that integrated four HISs commonly used in radiology department into a super-HIS called DDD. We adopted loose-coupling techniques in database integration and proposed a three-layer integration framework, including a data mashup layer, an aggregation service layer, and a result presentation layer. DDD enabled end-to-end tracking of the care each patient receives in radiology department, with detailed time stamps and contents of each care activity as well as rich information on patients, providers, and equipment. Furthermore, we proposed nine QoC metrics defined upon DDD: exam duration, technologist post-processing time, technologist TAT, radiologist TAT, and total TAT, which reflect the timeliness and efficiency of radiologic care; patient waiting time and patient TAT, which reflect efficiency and patient satisfaction; patient volume and exam volume, which reflect the workload distribution. All of these metrics measure QoC in radiology from different but complementary perspectives. DDD was deployed through a web portal, RQD, in MCA in the second quarter of 2016. Since then, it has been used extensively by clinicians, administrators, and researchers to monitor QoC, identify problem areas, and perform interventions to improve the quality of radiologic care. Specifically, we demonstrated, through four examples, that how users can use RQD information to understand the care workflow and performance. RQD may provide answers to some clinical practice related questions and concerns, and help identify opportunities for quality improvement. In addition, we showed a case study on a DDD-enabled intervention that effectively reduced the radiologist TAT for ObP. Specifically, our comparison between the pre- and post-intervention radiologist TAT showed that the intervention significantly eliminated ObP exams that take the residents longer than one hour to complete and cut the average TAT by more than half.

Several heuristic rules were adopted to handle the human errors that are inevitable in databases which require manual information entry. Over the course of the project, we have observed some issues related to discrepancies and inconsistencies of the data from human errors. The experienced radiologists and imaging informatics scientists from Mayo Clinic helped us understand the details of the radiology exam procedures and provided several examples that contain obvious input errors such as, wrong exam date in "ServiceTime", and improper check-in time (e.g., two days before exam date). Together, we developed heuristic rules to solve potential problems automatically for the HIS databases deployed in Mayo Clinic. Since the rules are applied to the time stamps collected from three commercial off-the-shelf HISs, the heuristic rules should be generalizable and applicable to other hospital systems. This has not yet been validated as the staff from other healthcare organizations may have different working conditions. We want to emphasize that even if the rules cannot be directly applied to other organizations, these rules could provide some guidelines for other practitioners looking to mitigate human errors as we did at Mayo Clinic.

The integration of the DIT database may limit the usage of DDD and RQD system since DIT is not a commercial system that is available for all facilities. DIT provides several types of accurate timestamps and exam information for calculation and analysis. We believe DIT can be replicated in other healthcare organizations with some (not substantial) effort as most such information can be acquired directly from the DICOM header in scanned images. For those hospital, clinic and healthcare facilities where DIT or a similar system has not yet been deployed, the practitioners could consider implementing a program to parse the necessary information from images. We have also received multiple inquiries about the DIT system since 2011 and currently, more than 20 members of the European Community of Medical Physicists have joined a fully collaborative effort facilitate broader use of DIT for addressing quality assurance issues.

As with other domain specific integrations, the domain experts, radiologists and imaging informatics scientists in Mayo Clinic, played an important role in the research and helped us in several ways: they helped us understand the radiology exam procedures and illustrate the timelines in all HIS as shown in Figure 4; they pointed out the potential errors and verified our heuristic rules; and they provided feedback to the defined metrics in terms of what they expect and how they want to compare metrics among different modalities, patient types and sites.

There are several future research directions we would like to pursue. First, we are continuously enriching our collection of QoC metrics. Second, advanced analytics can be performed on the rich datasets generated by DDD for QoC monitoring, abnormality detection, automatic alarming, and root cause identification, and to support other decision making in radiology.

Chapter 3

INTEGRATION OF SPARSE SINGULAR VECTOR DECOMPOSITION AND STATISTICAL PROCESS CONTROL FOR MONITORING AND CHANGE DETECTION OF HIGH-DIMENSIONAL DATA STREAMS

3.1 Introduction

Recent years have witnessed a significant increase in the use of wireless communication networks across private and public sectors. Mission-Critical Communication Networks (MCCNs) are those whose malfunction or failure will result in serious impact and even catastrophes (Baker and Hoglund, 2008). MCCNs are used widely in both civil and military settings. For example, public-safety first responders such as police officers, fire fighters, and paramedics use MCCNs to keep connected with each other and with the control center when responding to emergencies such as accidents, natural disasters, and terrorist attacks. Soldiers in a battle fields use MCCNs to communicate with each other and with the command center to acquire situational awareness and make tactical decisions. The nature of MCCNs puts an extremely high standard on the Quality of Service (QoS) these networks must provide. QoS refers to the performance of a communication network that is perceived by the users (International Telecommunication Union, 1993). A network with poor QoS delivers traffic data with delay, jitter, loss, or/and errors. QoS assurance for MCCNs is critically important for public safety, economic vitality, and national security. However, the existing approaches cannot directly be applied on data in MCCNs since their assumption of normality in monitored data cannot hold.

QoS assurance starts from monitoring and change/anomaly detection of network traffic data. This has been primary studied by the research community of communication networks in Electrical and Computer Engineering (ECE). A typical form of network traffic data is “packets”. A packet is a unit of data that is routed from a sender to a receiver in a network. A packet is typically structured to include a header and contents. The header includes meta-information about the packet such as sender and receiver IP addresses and protocol. Contents are the actual data such as text, audio, and video. The header of a packet is very small in size while contents can be large. The existing research falls into three major categories: The existing research falls into three major categories: deep packet inspection (DPI) (Yu *et al.*, 2006; Roesch *et al.*, 1999; Smith *et al.*, 2008; Cascarano *et al.*, 2011), active monitoring (AM) (Ciavattone *et al.*, 2003; Cáceres *et al.*, 1999; Paxson *et al.*, 1998; Almes *et al.*, 1999a,b), and passive monitoring (PM) (Fraleigh *et al.*, 2003; Ahmed *et al.*, 2005; Conway, 2002).

DPI examines packet contents, which has a major concern of privacy breach. AM and PM do not have this issue. AM works by injecting probing packets into the network and tracking these packets to detect QoS problems. A major drawback of AM is that it may disturb normal network operations. PM analyzes real packet data. The mainstay PM tools are relatively simple statistical methods, which do not suffice for MCCNs because of two reasons: First, network-wide monitoring, change detection, and fault diagnosis are needed because MCCNs are typically deployed to perform coordinated team work, while the existing PM methods focus on individual nodes, links, or sub-networks. Second, highly efficient algorithms are needed for real-time analytics of packet data in MCCNs that are temporally high-throughput and spatially densely-connected networks.

Packets that flow through sender-receiver pairs in a network are multivariate high-dimensional data streams (Stallings and T, 2013). Monitoring and change detection for high-dimensional data streams has been an important topic in Multivariate Statistical Process Control (MSPC) – a research area in Quality Engineering (QE). The existing approaches integrate various variable selection (VS) or sparse learning techniques with classic MSPC control charts to address high-dimensionality (Bersimis *et al.*, 2007; Zou and Qiu, 2009; Capizzi and Masarotto, 2011; Wang and Jiang, 2009; Jiang *et al.*, 2012). Also available are two-step approaches that first construct a statistic for each individual data stream and then combine the statistics in a way to achieve global monitoring (Tartakovsky and Veeravalli, 2008; Mei, 2010; Zou *et al.*, 2015; Zhang, 2002). While these existing approaches are based on sophisticated statistics and thus being potentially useful for complementing PM in monitoring packet data of MCCNs, their direct application is impractical because of the following reasons: First, most existing approaches assume Gaussian data; the two-step approaches are optimal when the data streams are independent. However, packet data in MCCNs are non-Gaussian and have an inherent correlation structure. Second, most existing approaches focus on mean shift detection and assume the shift to be a step change. However, changes in MCCNs are mean shifts coupled with variance and covariance changes. Also, the temporal shape of a change in MCCNs can be more than just a step, but also include trends and oscillating/trembling patterns, with each shape corresponding to a different root cause. The ability for differentiating the temporal shapes of changes is important for root cause identification and QoS improvement. Third, real-time monitoring of MCCNs needs highly efficient analytics algorithms. This standard can be hardly met by the existing approaches.

Noting the gap that neither the QoS research on communication networks nor

MSPC research on high-dimensional data streams has been able to provide an effective tool for MCCN packet data monitoring, we propose a new method called Sparse Singular Value Decomposition (SSVD)-MSPC to fill in this gap. SSVD-MSPC has the following novel features:

- By developing a monitoring statistic based on SVD, SSVD-MSPC provides a non-parametric approach, which can accommodate the special packet data distribution and correlation structure.
- By integrating SVD and sparse learning, SSVD-MSPC addresses the challenge of accurate fault identification from high-dimensional stochastic packet streams.
- SSVD-MSPC provides three key capabilities toward QoS improvement of MCCNs, including 1) monitoring, i.e., analyzing packet data and detecting changes from the in-control distribution; 2) fault identification, i.e., locating the faulty sender, receiver, or links that are responsible for the change; and 3) fault characterization, i.e., estimating the temporal shape of the change/fault.
- SSVD-MSPC is computationally efficient and therefore suites the need for network-wide real-time monitoring, fault identification and characterization of temporally high-throughput and spatially densely-connected MCCNs.

3.2 Literature Review

This research intersects with two existing areas: QoS research on communication networks, which has been primarily studied in ECE; MSPC research on high-dimensional data streams, which has been primarily studied in QE. Next, we will review the works in each area and discuss their limitations that justify the need for new methodological development.

3.2.1 QoS Research on Communication Networks

QoS assurance starts from monitoring and change/anomaly detection of network traffic data. This has been primary studied by the research community of communication networks in ECE. A typical form of network traffic data is “packets”. A packet is a unit of data that is routed from a sender to a receiver in a network. A packet is typically structured to include a header and contents. The header includes meta-information about the packet such as sender and receiver IP addresses and protocol. Contents are the actual data such as text, audio, and video. The header of a packet is very small in size while contents can be large. The existing research falls into three major categories: deep packet inspection (DPI) (Yu *et al.*, 2006; Roesch *et al.*, 1999; Smith *et al.*, 2008; Cascarano *et al.*, 2011), active monitoring (AM) (Ciavattone *et al.*, 2003; Cáceres *et al.*, 1999; Paxson *et al.*, 1998; Almes *et al.*, 1999a,b), and passive monitoring (PM) (Fraleigh *et al.*, 2003; Ahmed *et al.*, 2005; Conway, 2002).

DPI examines the contents of packets passing through a so-called an inspection point within a network, and searches for anything out of the norm. While DPI can be used to detect QoS problems, its major utility is to ensure network security by detecting instructions, viruses, spams, and non-compliance of contents with regulation. Various DPI systems and techniques have been developed. For example, SNORT (Roesch *et al.*, 1999) is a well-known open-source system that can detect various types of worms, attacks, and probes using protocol analysis, and content searching and matching. (Smith *et al.*, 2008) proposed a DPI technique that uses regular expression with extended finite automata. Focusing on QoS, (Cascarano *et al.*, 2011) proposed and validated optimizations for DPI techniques to accelerate network monitoring and traffic classification on high-speed networks. There are several drawbacks in using

DPI for QoS: First, DPI often requires costly dedicated devices to track, unpack, and analyze real-time packets. Second, DPI can be time-consuming especially with large-sized packet contents (e.g., audio, video), which makes it unsuitable for real-time QoS monitoring. Third, because DPI examines packet contents, there is a profound concern on privacy. Fourth, due to the privacy concern, more and more network protocols such as HTTPS, SFTP and SSL are designed to protect private contents from being examined by DPI. To overcome these limitations, other methods have been developed for QoS assurance, which fall in two categories, AM and PM.

AM works by injecting "probing packets" into the network and tracking these packets to assess QoS. Various methods for AM have been developed. For example, (Ciavattone *et al.*, 2003) proposed a network-wide AM system that is operated on a tier 1 IP backbone and can monitor several QoS metrics such as packet delay, loss, traceroute, delay variation, and reordered or out-of-order packets. (Cáceres *et al.*, 1999) proposed an algorithm to estimate packet loss rates on individual links based on losses observed by multicast receivers via maximum likelihood estimators. (Paxson *et al.*, 1998) proposed a measurement infrastructure for National Internet Measurement Infrastructure (NIMI) for assessing the performance of an Internet path, including one-way and round-trip loss and delay, available bandwidth, and routing stability. Almes et. al. proposed QoS metrics for IP performance metrics (IPPM) in several Request for Comments documents (Almes *et al.*, 1999a,b). The major drawbacks of AM are fairly obvious: Because it injects additional traffic into the network, it disturbs the network's normal operations. This makes it unfit for MCCNs for which even minimum disturbance may cause catastrophe. Also, QoS of the probing packets may not precisely reflect the behaviors of real packets in the network.

Unlike AM, PM tracks and analyzes real packet data in the network. Unlike DPI, PM does not inspect packet contents but uses statistical and visualization tools to analyze network traffic behaviors such as packet passing rate, inter-arrival times, delay, loss, and queue size. (Fraleigh *et al.*, 2003) proposed a PM system to analyze packet queuing and transmission behaviors. (Ahmed *et al.*, 2005) implemented a policy-based management system based on real-time traffic flow measurements to achieve dynamic QoS adaptation for multimedia applications. (Conway, 2002) developed a PM method for monitoring speech quality in live (i.e. in progress) VoIP calls. The existing PM research is primarily conducted in the communication network research communities of ECE. Relatively simple statistical methods have been used, which leave the following areas for improvement: First, network-wide monitoring, change detection, and fault diagnosis are much needed, while existing research primarily focuses on individual nodes, links, or small local sub-networks. This capability is particularly important for MCCNs because of an MCCN is typically deployed to perform coordinated team work. Second, real-time analysis is important, which poses a high standard on the processing speed of the analytic method. This is particularly challenging for MCCNs that are temporally high-throughput and spatially densely-connected networks.

3.2.2 MSPC Research on High-Dimensional Data Streams

Packets that flow through sender-receiver pairs in a network are multivariate high-dimensional data streams. Monitoring and change detection for high-dimensional data streams has been an important topic in MSPC in recent years. One category of approaches is to integrate VS techniques (a.k.a. sparse learning) with classic MSPC (Bersimis *et al.*, 2007) – referred to as VS-MSPC hereafter – under the assumption that changes only occur in a small subset of all the data streams. For example, (Zou and Qiu, 2009) proposed a method that integrates adaptive LASSO (Zou, 2006) with EWMA control charts for mean shift detection. (Capizzi and Masarotto, 2011) proposed a combination of Least Angle Regression (LAR) (Efron *et al.*, 2004) with a multivariate EWMA control chart for detection of shifts in both the means and the total variability. (Wang and Jiang, 2009) developed a Shewhart-type multivariate control chart that uses a forward variable selection method to select suspicious variables. This approach was further extended to a multivariate EWMA procedure (Jiang *et al.*, 2012). Another category of approaches first constructs a statistic for each individual data stream and then combines the statistics in a way to achieve global monitoring. Call these approaches "combined individuals" hereafter. For example, (Tartakovsky and Veeravalli, 2008) proposed a T_{max} approach that constructs a CUSUM statistic for each data stream and then monitors the maximum of the CUSUM statistics. T_{max} assumes that changes occur in exactly one out of all p data streams. Alternatively, (Mei, 2010) proposed a T_{sum} approach that monitors the sum of the individual CUSUM statistics and showed through simulations that T_{max} is more effective than T_{sum} when changes occurs in only a few data streams, but is outperformed by T_{sum} when changes occurs in a moderate to large number of data streams. (Zou *et al.*, 2015) developed a control chart that balances the detection abilities of T_{max} and T_{sum} using a powerful goodness-of-fit test proposed in (Zhang, 2002).

As promising as the afore-reviewed approaches seem for monitoring of packet data in MCCNs, they are not directly applicable because of the following reasons: 1) Most existing approaches assume that the data streams are multivariate Gaussian. The "combined individuals" approaches are optimal when the data streams are independent. However, network-wide packet data are not Gaussian and are inherently correlated. Relevant to QoS, each packet sent from a sender i to a receiver j can be represented by a binary variable, x_{ij}. $x_i j = 0$ or 1 means that the packet is lost or through. The number of through packets during a time interval, y_{ij}, characterizes the QoS, and is not Gaussian. Also, most MCCNs are multicast networks in which each node can send packets to multiple or all other nodes simultaneously as receivers. This naturally creates correlations between packet streams. 2) Most existing approaches focus on mean shift detection and assume the shift to be a step change. However, changes in MCCNs are mean shifts coupled with variance and covariance changes. More in-depth discussion of this property will be provided in Section 3. Furthermore, the temporal shape of a change can be more than just a step, but also include trends and oscillating/trembling patterns. Each shape may correspond to a different root cause: a step change can be caused by a device failure or congestion; a trend change can be due to battery wear out or communication gradually out of range; an oscillating change can be due to environmental interference. It is important for a packet monitoring and change detection method to be able to differentiate the different shapes for effective root cause diagnosis. This ability is lacking in the existing MSPC approaches for high-dimensional data stream monitoring. 3) MCCNs generate high-throughput packet data; the multicast communication creates high-dimensional data streams. Real-time monitoring of such temporally-spatially challenging networks needs highly efficient analytics algorithms. This standard can hardly be satisfied by the existing approaches.

3.3 SSVD-MSPC for Monitoring, Fault Identification and Characterization of MCCNs

3.3.1 Statistical Structure and Properties of Packet Data in MCCNs

Consider an MCCN with N users/nodes. A packet can be successfully sent from node i (called a sender) to node j (called a receiver) if and only if three conditions are met: 1) the packet is successfully sent by the sender, 2) successfully travels through the link between the sender and the receiver, and 3) is successfully received by the receiver. Let p_i^S, p_{i-j}^L, and p_j^R be the success probabilities for the sender, link, and receiver, respectively. Then, the number of through packets during a time interval Δt, x_{ij}, follows a Binomial distribution, i.e., $x_{ij} \sim Binomial(n, p_{ij})$, where $p_{ij} = p_i^S \times p_{i-j}^L \times p_j^R$ and n is the intended number of packets sent from i to j during Δt. n can be obtained by multiplying Δt by the packet firing rate that is known from the network design. For example, this rate is roughly 50 packet-per-second for voice transmission in mobile networks (Cisco System, Inc., 2016). Under the normal operating condition (i.e., the "in-control" condition in MSPC terminology), p_i^S, p_{i-j}^L, and p_j^R are high. The higher the QoS standard of a network, the greater the p_i^S, p_{i-j}^L, and p_j^R.

Furthermore, let $\mathcal{S}$ and $\mathcal{R}$ denote the collections of senders and receivers in an MCCN, respectively. $\mathcal{S}$ and $\mathcal{R}$ have the following properties: (a) Each node can send and receive packets, which means that a sender can also be a receiver. (b) A sender can send packets to more than one receiver at a time, which results in a multicast network (Williamson, 2000). (c) A receiver can receive packets from more than one sender. Let $\mathbf{x}$ be the packet data in all sender-receiver pairs of a network, $\mathbf{x} = \{x_{ij} : i \in \mathcal{S}, j \in \mathcal{R}\}$. $\mathbf{x}$ follows a multivariate Binomial distribution with element-wise means and variances given in Equation 3.1 and a special covariance structure in Equation 3.2 as a results of the aforementioned properties (b)-(c)):

$$E\left[x_{ij}\right] = np_i^S p_{i-j}^L p_j^R, var\left(x_{ij}\right) = np_i^S p_{i-j}^L p_j^R \left(1 - p_i^S p_{i-j}^L p_j^R\right) \tag{3.1}$$

$$cov\left(x_{ij}, x_{kh}\right) =$$

$$\begin{cases} 0 & \text{if } i \neq k \text{ and } j \neq h \\ & (x_{ij} \text{ and } x_{kh} \text{ do not share the same sender or receiver}) \\ n\left(1 - p_i^S\right) p_i^S p_j^R p_h^R p_{i-j}^L p_{k-h}^L & \text{if } i = k \text{ and } j \neq h \\ & (x_{ij} \text{ and } x_{kh} \text{ share the same sender but not receiver}) \\ n\left(1 - p_j^R\right) p_j^R p_i^S p_k^S p_{i-j}^L p_{k-h}^L & \text{if } i \neq k \text{ and } j = h \\ & (x_{ij} \text{ and } x_{kh} \text{ share the same receiver but not sender}) \end{cases} \tag{3.2}$$

Derivations to get Equation 3.2 are skipped.

Three types of faults can occur in a network: sender, receiver, and link faults. Their respective definitions and impacts on the distribution of $\mathbf{x}$ are summarized in Table 3. The root causes could be malfunction of physical communication devices, which can cause sender or receiver faults; congestion of communication channels, which can cause link faults; environmental interferences such as severe weather conditions and geographic blocks, which can cause sender, receiver, or link faults depending on if the interferences affect a sender, receiver, or a link, respectively. Note that Table 3 only shows the impact of faults on the means, variances, and covariances (i.e., the first and second moments) of the multivariate Binomial distribution of $\mathbf{x}$, while the faults can impact higher-order moments.

Furthermore, even within the same type of fault, there are subtypes corresponding to different temporal shapes of the fault. For example, Figure 10 shows three different

Table 3: Definitions and impacts of sender, receiver, and link faults (↓ and ↑ represent decrease and increase from the in-control parameters, respectively)

	Definition	Impact on $\mathbf{x}$ distribution
Sender fault	When sender i is faulty, $p_i^S \downarrow$.	$E\left[x_{ij}\right] \downarrow, var\left(x_{ij}\right) \uparrow,$ $cov\left(x_{ij}, x_{ih}\right) \uparrow$
Receiver fault	When receiver j is faulty, $p_j^R \downarrow$.	$E\left[x_{ij}\right] \downarrow, var\left(x_{ij}\right) \uparrow,$ $cov\left(x_{ij}, x_{kj}\right) \uparrow$
Link fault	When link $i-j$ is faulty, $p_{i-j} \downarrow$.	$E\left[x_{ij}\right] \downarrow, var\left(x_{ij}\right) \uparrow,$ $cov\left(x_{ij}, x_{ih}\right) \downarrow, cov\left(x_{ij}, x_{kj}\right) \downarrow$

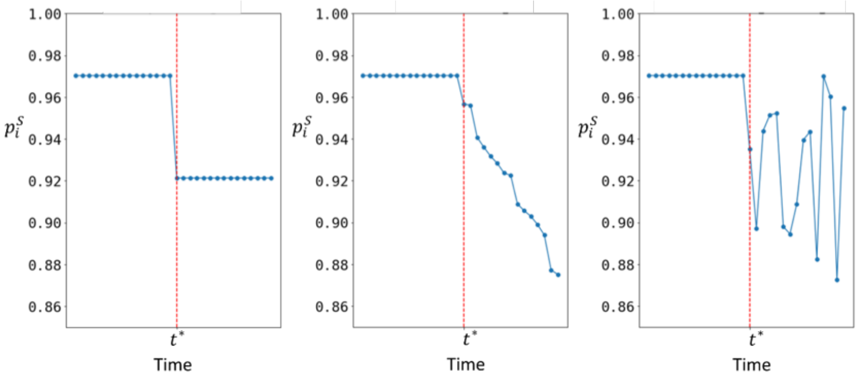


Figure 10: Step (left), trend (middle), and oscillating (right) changes of the fault in MCCNs (t^* indicates the change point).

shapes that can occur for a sender fault: a step change can be caused by a device failure or congestion; a trend change can be due to battery wear out or communication gradually out of range; an oscillating change can be due to environmental interference.

3.3.2 Development of SSVD-MSPC

For QoS assurance in MCCNs, it is important to develop a method to accomplish three tasks: 1) monitoring, which is to analyze packet data $\mathbf{x}$ and fire an alarm when the distribution of $\mathbf{x}$ changes from its in-control distribution; 2) fault identification, which is to find the faulty sender, receiver, or link responsible for the change in 1); 3) fault characterization, which is to estimate the temporal shape of the change/fault in 2).

To accomplish the monitoring task, we can build a control chart on $\mathbf{x}$. The true distribution of $\mathbf{x}$ is multivariate Binomial. Although we may approximate $\mathbf{x}$ by multivariate Gaussian, the challenge is that changes in the distribution due to faults are mean shifts coupled by variance and covariance changes (see Table 3). To our best knowledge, there is no control chart designed for detecting such complicatedly coupled changes in a multivariate Gaussian distribution. This leaves us a few options: One is to develop such a control chart under the Gaussian approximation to Binomial; another is to develop a control chart for multivariate Binomial data; the third option is to develop a non-parametric method. We choose the third option due to its analytical simplicity, computational efficiency, and the ease for facilitating fault identification and characterization after the monitoring stage.

Specifically, let $\mathbf{X}$ be a $q \times m$ matrix that contains the packet data within a monitoring time window. m is the window size. q is the number of sender-receiver pairs in the network. Let $x_{ij}(t)$ be an element of $\mathbf{X}$ that corresponds to the sender-receiver pair (i, j) at time t, $t = 1, \ldots, m$. Let $\mathcal{F}$ denote the set of sender-receiver pairs affected by a fault. For a pair $(i, j) \notin \mathcal{F}$, $E[x_{ij}(t)] = np_0$, where $p_0 = p_0^S \times p_0^L \times p_0^R$. p_0^S, p_0^L, and p_0^R are in-control success probabilities for senders, links, and receivers, which are known from the network design. For a pair $(i, j) \in \mathcal{F}$, $E[x_{ij}(t)] = n(p_0 - \delta(t))$, where $\delta(t) \geq 0$ for $t = 1, \ldots, m$ and there exists at least one t for which $\delta(t) > 0$. $(\delta(1), \ldots, \delta(m))^T$ can be of different shapes with examples given in Figure 10. Next, focus on $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$, which measures the deviation of the expected packet data from that under the in-control condition. $\mathbf{1}_{q\times m}$ is a matrix of all ones. We show in

Proposition 1 that the SVD of $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$ has only one non-zero singular value, i.e., the first singular value. Also, the first left- and right-singular vectors capture the faulty sender-receiver pairs and the temporal shape of the fault, respectively. A brief introduction to SVD and proof of Proposition 3.1 are provided in Appendix I and II.

Proposition 3.1: Let s_k, $\mathbf{u}_k$, $\mathbf{v}_k$ be the k-th singular value, left- and right-singular vectors for the SVD of $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$. Let $|\mathcal{F}|$ denote the cardinality of $\mathcal{F}$. Then, (a) $s_1 > 0$ and $s_1 = \sqrt{\sum_{t=1}^{m} [n\delta(t)]^2} \times |\mathcal{F}|$; $s_k = 0$ for $k > 1$; (b) $\mathbf{u}_1 = u_{1,ij} : i \in \mathcal{S}, j \in \mathcal{R}$ where $u_{1,ij} = \frac{1}{\sqrt{|\mathcal{F}|}}$ if $(i,j) \in \mathcal{F}$ and $u_{1,ij} = 0$ otherwise; (c) $v_1 = \frac{-1}{\sqrt{\sum_{t=1}^{m} [\delta(t)]^2}}(\delta(1), \ldots, \delta(m))^T$.

Proposition 3.1 suggests that we may use s_1 as a monitoring statistic, use $\mathbf{u}_1$ to locate faculty sender-receiver pairs, which further allow us to identify which sender, receiver, or links are at fault, and use $\mathbf{v}_1$ to characterize the temporal shape of the fault. A practical challenge, however, is that the packet data we have is $\mathbf{X}$ not $E[\mathbf{X}]$. To account for the stochastic nature of $\mathbf{X}$, we propose the following modifications on the results in Proposition 3.1 to accomplish the three tasks of monitoring, fault identification, and fault characterization. By an abuse of notation, we re-use s_1, $\mathbf{u}_1$, and $\mathbf{v}_1$ to denote the first singular value, left- and right-singular vectors of $\mathbf{X} - np_0\mathbf{1}_{q\times m}$.

3.3.2.1 Monitoring

Due to the stochastic nature of $\mathbf{X}$, s_1 is not zero even under the in-control condition. Therefore, we cannot use zero as the upper control limit (UCL) for s_1 to alarm for out-of-control conditions. Note that there is no lower control limit (LCL) since $s_1 \geq 0$ by definition of SVD. Also, s_1 does not follow any known parametric distribution, because the SVD is performed on data from a multivariate Binomial distribution. We propose a non-parametric approach. Specifically, we develop a simulation platform that takes network design information as input, including a sender-receiver correspondence matrix, $\mathbf{C}$, and the in-control success probabilities for the senders, links, and receivers, p_0^S, p_0^L, and p_0^R, and generate packet data $\mathbf{X}$ under the in-control condition. Then, SVD will be performed on $\mathbf{X} - np_0\mathbf{1}_{q\times m}$ to obtain s_1. Because the simulation is performed offline, we can run it long enough to collect a large number of samples for s_1, which will further allow us to obtain the empirical in-control distribution of s_1 and use the $(1-\alpha)$-th percentile of the distribution as the UCL. α is a pre-selected Type I error probability. Denote the UCL by $UCL_{s_1}(\alpha)$. Next, we discuss the details of the simulation platform.

From the sender-receiver correspondence matrix, $\mathbf{C}$, we can know the set of receivers for each sender i. Let $\mathbf{R}_i$ denote the receiver set and $|\mathbf{R}_i|$ be the number of receivers. For each intended packet that is to be sent from sender i to its receivers, the first step is to determine if the packet is successfully sent out. This is done by generating a Bernoulli sample with probability p_0^S. If the sample is zero, label the packet as "failed". Otherwise, make $|\mathbf{R}_i|$ copies of the packet and assign one copy to each link $i-j$, $j \in \mathbf{R}_i$. This follows from the design of multicast networks. The second step is to determine if the packet (copy) assigned to link $i-j$ successfully travels through the link and is successfully received by receiver j. To achieve this, sample from a Bernoulli distribution with probability p_0^L. If the sample is zero, label the packet as "failed". Otherwise, sample from a Bernoulli distribution with probability p_0^R. If the sample is zero, label the packet as "failed". Otherwise, label it as "through". Applying this two-step procedure to n intended packets, we can obtain the number of through packets for each sender-receiver pair of the network, which composes one column of $\mathbf{X}$. Other columns of $\mathbf{X}$ can be generated in the same way.

For online monitoring of real-time generated packet data in an MCCN, $\mathbf{X}$, we can compute the first singular value s_1 of $\mathbf{X} - np_0\mathbf{1}_{q\times m}$ using the "dgesdd" routine from Linear Algebra PACKage (LAPACK) (Anderson *et al.*, 1999). "dgesdd" is a very efficient algorithm for computing singular values of a matrix without having to carry out the complete SVD (Blackford, 1999). Then, the s_1 is compared with the $UCL_{s_1}(\alpha)$ obtained from offline simulation. If $s_1 > UCL_{s_1}(\alpha)$, it is an indication for an out-of-control condition or a potential fault happening in the network. This will trigger the subsequent fault identification and characterization.

3.3.2.2 Fault Identification

Proposition 3.1 implies that we may use the non-zero elements of $\mathbf{u}_1$ to locate faulty sender-receiver pairs. However, because the SVD is performed on $\mathbf{X} - np_0\mathbf{1}_{q\times m}$ not $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$, no element of $\mathbf{u}_1$ will be exactly zero. To overcome this problem, we propose an SSVD to obtain a sparse estimator for $\mathbf{u}_1$. Recall that the regular SVD (see Appendix I for an introduction) estimates $\mathbf{u}_1$ by solving the following optimization problem:

$$(\hat{s}_1, \hat{\mathbf{u}}_1, \hat{\mathbf{u}}_1) = \underset{s_1,\mathbf{u}_1,\mathbf{v}_1}{\arg\min} \| (\mathbf{X} - np_0\mathbf{1}_{q\times m}) - s_1\mathbf{u}_1\mathbf{v}_1^T \|_F^2,$$

$$s.t. \, \|\mathbf{u}_1\|^2 = 1, \|\mathbf{v}_1\|^2 = 1, s_1 \geq 0$$

where $\| \cdot \|_F^2$ is the squared-Frobenius norm of a matrix and $\| \cdot \|_2$ is the l_2-norm of a vector. To impose sparsity on the estimation for $\mathbf{u}_1$, SSVD solves the following l_1-penalized optimization:

$$(\tilde{\mathbf{u}}_1^s, \tilde{\mathbf{v}}_1) = \underset{\mathbf{u}_1^s,\mathbf{v}_1}{\arg\min} \| (\mathbf{X} - np_0\mathbf{1}_{q\times m}) - \mathbf{u}_1^s\mathbf{v}_1^T \|_F^2 + \lambda\|\mathbf{u}_1^s\|_1 \tag{3.3}$$

where $\mathbf{u}_1^s = s_1\mathbf{u}_1$, $\|\cdot\|_1$ is the l_1-norm, and λ is a tuning parameter. Adding the l_1-penalty to $\mathbf{u}_1^s$ not $\mathbf{u}_1$ makes the optimization easier to solve because it relaxes the unity constraint on $\mathbf{u}_1$. Equation 3.3 also relaxes the unity constraint on $\mathbf{v}_1$ because our purpose here is not to perform an SVD but use the non-zero elements in $\hat{\mathbf{u}}_1$ to locate faulty sender-receiver pairs.

The optimization problem in Equation 3.3 can be solved by alternating between two sub-problems until convergence. That is, given $\mathbf{u}_1^s$, Equation 3.3 becomes

$$\mathbf{v}_1^* = \arg\min_{\mathbf{v}_1} \|\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right) - \mathbf{u}_1^s\mathbf{v}_1^T\|_F^2 \tag{3.4}$$

which can be solved analytically, i.e., $\mathbf{v}_1^* = \left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)^T\mathbf{u}_1^s$. Given $\mathbf{u}_1^s$, Equation 3.3 becomes

$$\begin{aligned}\mathbf{u}_1^{s*} &= \arg\min_{\mathbf{u}_1^s} \|\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right) - \mathbf{u}_1^s\mathbf{v}_1^T\|_F^2 + \lambda\|\mathbf{u}_1^s\|_1 \\ &= \arg\min_{\mathbf{u}_1^s} \|\mathbf{X} - np_0\mathbf{1}_{q\times m}\|_F^2 + \sum_{i=1}^{q}\left\{\left(u_{1i}^s\right)^2 - 2u_{1i}^s(\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i + \lambda|u_{1i}^s|\right\}\end{aligned}$$

where u_{1i}^s is the i-the element of $\mathbf{u}_1$. This suggests that we can minimize each u_{1i}^s separately by solving q optimization problems in the form of Equation 3.5:

$$u_{1i}^{s\,*} = \arg\min_{u_{1i}^s} \left(u_{1i}^s\right)^2 - 2u_{1i}^s(\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i + \lambda|u_{1i}^s| \tag{3.5}$$

$i = 1, \ldots, q$. Furthermore, Proposition 3.2 shows that Equation 3.5 has a close-form solution, the proof of which uses the KKT condition and is skipped due to space limit.

Proposition 3.2: The minimizer of the optimization problem in Equation 3.5 is $u_{1i}^{s\,*} = sign\left\{(\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i\right\}\left\{|(\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i| - 0.5\lambda\right\}_+$. That is,

$$u_{1i}^{s\,*} = \begin{cases} (\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i - 0.5\lambda, & \text{if } (\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i \geq 0.5\lambda \\ (\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i + 0.5\lambda, & \text{if } (\left(\mathbf{X} - np_0\mathbf{1}_{q\times m}\right)\mathbf{v}_1)_i < -0.5\lambda \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

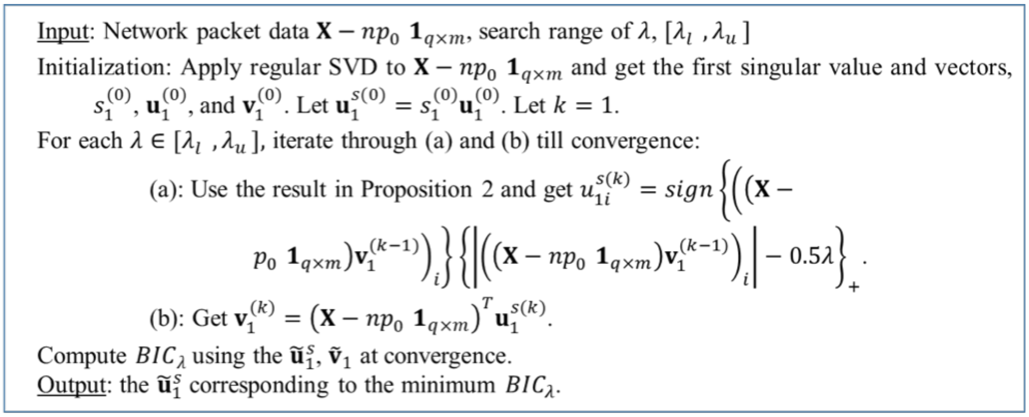
Next, we discuss the selection of the tuning parameter λ. This problem is known as "model selection" in statistics. Common model selection criteria include BIC, AIC, and cross validation. In this paper, because our SSVD formulation is similar to LASSO, we follow the suggestion in (Zou *et al.*, 2007) and adopt BIC. The BIC for our model in 3.3 is:

$$BIC_\lambda = \frac{\| (\mathbf{X} - np_0 \mathbf{1}_{q \times m}) - \tilde{\mathbf{u}}_1^s \tilde{\mathbf{v}}_1^T \|_F^2}{qm\sigma_{OLS}^2} + \hat{df}_\lambda \frac{log(qm)}{qm} \tag{3.7}$$

where $\hat{df}_\lambda$ is the degree of freedom of $\tilde{\mathbf{u}}_1^s$, i.e., the number of non-zero elements in $\tilde{\mathbf{u}}_1^s$. σ_{OLS}^2 is the ordinary least squares (OLS) estimator for the error variance σ^2 as in the regression model $\mathbf{Y} = (\tilde{\mathbf{u}}_1^s)^T \left(\mathbf{I}_q \otimes \tilde{\mathbf{v}}_1^T\right) + \epsilon$, $\epsilon \sim N(0, \sigma^2 \mathbf{I}_{qm})$, where $\mathbf{Y}$ is the concatenation of all row vectors from matrix $(\mathbf{X} - np_0 \mathbf{1}_{q \times m})$ and $\otimes$ denotes the Kronecker product. Finally, we summarize the algorithm for solving SSVD in Figure 11. Note that although the algorithm also produces an estimate for $\mathbf{v}_1$, it is not useful for fault identification. Therefore, we may only output $\tilde{\mathbf{u}}_1^s$ in which the non-zero elements correspond to faulty sender-receiver pairs in the network. The algorithm is very efficient because it iterates between two sub-optimization problems that both can be solved analytically. Also, the algorithm is guaranteed to converge because the sub-optimization problems are both convex.

3.3.2.3 Fault Characterization

Once the faulty sender-receiver pairs are identified by the SSVD algorithm, we can further find out if those pairs share the same sender (a sender fault) or the same receiver (a receiver fault), or are a group of individual links. Fault characterization aims to estimate the temporal shape of the fault. To achieve this, we transform $\tilde{\mathbf{u}}_1^s$ estimated by the SSVD algorithm into a 0/1 indicator vector $\mathbf{u}_1^{0/1}$ by keeping the



```
Input: Network packet data $\mathbf{X} - np_0\ \mathbf{1}_{q\times m}$, search range of $\lambda$, $[\lambda_l\ , \lambda_u\ ]$
Initialization: Apply regular SVD to $\mathbf{X} - np_0\ \mathbf{1}_{q\times m}$ and get the first singular value and vectors,
    $s_1^{(0)}$, $\mathbf{u}_1^{(0)}$, and $\mathbf{v}_1^{(0)}$. Let $\mathbf{u}_1^{s(0)} = s_1^{(0)}\mathbf{u}_1^{(0)}$. Let $k = 1$.
For each $\lambda \in [\lambda_l\ , \lambda_u\ ]$, iterate through (a) and (b) till convergence:
        (a): Use the result in Proposition 2 and get $u_{1i}^{s(k)} = sign\left\{\left(\left(\mathbf{X} - p_0\ \mathbf{1}_{q\times m}\right)\mathbf{v}_1^{(k-1)}\right)_i\right\}\left\{\left|\left(\left(\mathbf{X} - np_0\ \mathbf{1}_{q\times m}\right)\mathbf{v}_1^{(k-1)}\right)_i\right| - 0.5\lambda\right\}_+$.
        (b): Get $\mathbf{v}_1^{(k)} = \left(\mathbf{X} - np_0\ \mathbf{1}_{q\times m}\right)^T\mathbf{u}_1^{s(k)}$.
Compute $BIC_\lambda$ using the $\tilde{\mathbf{u}}_1^s$, $\tilde{\mathbf{v}}_1$ at convergence.
Output: the $\tilde{\mathbf{u}}_1^s$ corresponding to the minimum $BIC_\lambda$.
```

Figure 11: The SSVD Algorithm for fault identification

zero elements and changing the nonzero elements to one. Then, we insert $\mathbf{u}_1^{0/1}$ into Equation 3.4 and solve for $\mathbf{v}_1$, i.e., $\check{\mathbf{v}}_1 = (\mathbf{X} - np_0\mathbf{1}_{q\times m})\,\mathbf{u}_1^{0/1}$. It is straightforward to prove that $\check{\mathbf{v}}_1$ is the average temporal profile of the sender-receiver pairs corresponding to elements ones in $\mathbf{u}_1^{0/1}$, i.e., the faulty pairs identified by SSVD. Furthermore, we can visually inspect the average temporal profile or use more rigorous hypothesis testing to characterize the shape of the profile (e.g., step, trend, oscillating functions of time).

Finally in this section, we summarize the SSVD-MSPC method for monitoring, fault identification, and fault characterization of packet data of MCCNs in Figure 12.

3.4 Case Studies on MCCN

Our industrial collaborator is a company that performs testing of MCCNs before they are deployed in the field. Because of the sensitive nature, we cannot use the real data generated from the testing, but can simulate data based on knowledge about the design and configuration of the MCCNs. There are three sizes of MCCNs: small, medium, and large. A medium-sized MCCN typically includes $10 - 20$ nodes and is

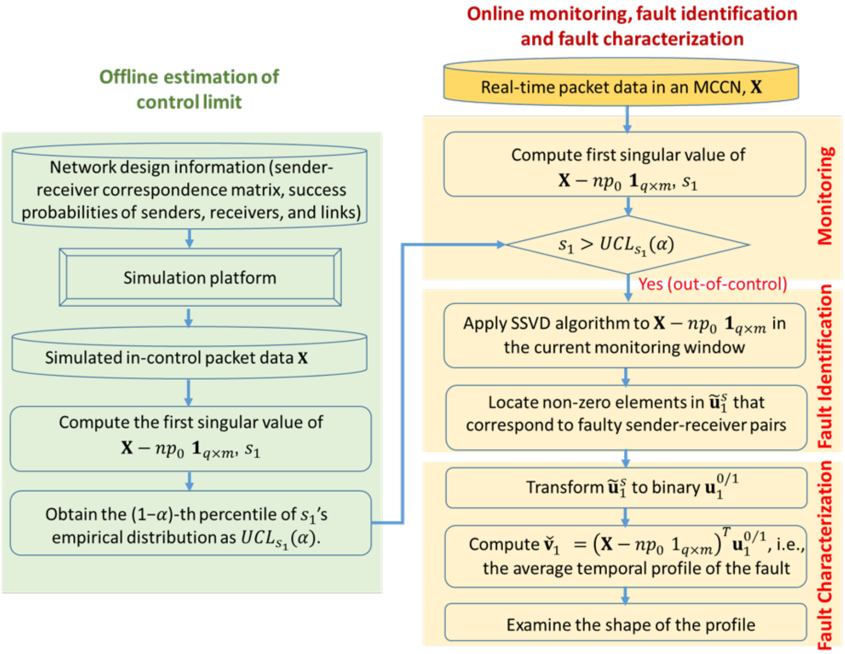


Figure 12: SSVD-MSPC for monitoring, fault identification, and fault characterization of packet data in MCCNs

the most common. We consider a 15-node medium-sized MCCN in our experiments. Because each node can be a sender and receiver, a 15-node MCCN includes 210 sender-receiver pairs for which the packet data needs to be monitored. We also consider a small 5-node MCCN that includes 20 sender-receiver pairs and a large 50-node MCCN that includes 2450 sender-receiver pairs. Each node is a mobile device that can transmit voice. The transmission rate is 50 packets per second, i.e., $n = 50$. Furthermore, according to the network design, the in-control success probabilities for senders, links, and receivers are $p_0^S = p_0^L = p_0^R = 0.97$. Based on the aforementioned parameters, we can obtain the control limit $UCL_{s_1}(\alpha)$ offline by following the steps on the left side of Figure 12. Specifically, $\alpha = 0.002$ is used in our experiments, which corresponds to an in-control average run length (ARL_0) of 500. Consequently, we get$UCL_{s_1}(0.002) = 11.17$, 24.56, and 66.69 for the MCCNs of 5, 15, and 50 nodes, respectively, under a monitoring window of $m = 10$ seconds.

3.4.1 Performance of SSVD-MSPC in monitoring MCCNs

We compare the performance of SSVD-MSPC with the method in (Zou and Qiu, 2009), which integrated adaptive LASSO with MSPC, referred to as LASSO-MSPC hereafter. LASSO-MSPC is a best-known method in modern MSPC literature for monitoring high-dimensional data streams. Both SSVD-MSPC and LASSO-MSPC are implemented using the R programming language. More specifically, the "lars" package in R is used to find the transition points needed to compute the test statistic in LASSO-MSPC; the "base" package is used to compute the test statistic (i.e., the first singular value) in SSVD-MSPC.

Because LASSO-MSPC was developed to detect step changes, we focus on step changes in this experiment. Specifically, packet data is generated with an out-of-control success probability of $0.97 - \Delta p$ for whichever sender, receiver, or link at fault. We set $\Delta p = 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04$ in order to assess the performance of SSVD-MSPC over a wide range of change/fault magnitudes. We focus on sender and link faults, because receiver faults impact packet data distribution in the same way as sender faults (see Table 3) and our experiments also showed similar performance in detecting receiver faults to sender faults by both SSVD-MSPC and LASSO-MSPC. For sender faults, we randomly select one sender in the MCCN to be faulty. Since a sender can send packets to all other nodes in the network, a sender fault will introduce changes to 4, 14, and 49 sender-receiver pairs in the 5, 15, and 50-node MCCNs, respectively. For link faults, we randomly select a subset of sender-receiver pairs

Table 4: Average CPU time (seconds) of fault/change detection

(a) sender fault

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	< 0.001	0.004	< 0.001	0.216	0.001	-
0.09	< 0.001	0.004	< 0.001	0.218	0.001	-
0.08	< 0.001	0.004	< 0.001	0.210	0.001	-
0.07	< 0.001	0.004	< 0.001	0.215	0.001	-
0.06	< 0.001	0.004	< 0.001	0.216	0.001	-
0.05	< 0.001	0.004	< 0.001	0.215	0.001	-
0.04	< 0.001	0.004	< 0.001	0.212	0.001	-

(b) link fault

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	< 0.001	0.004	< 0.001	0.258	0.001 -	
0.09	< 0.001	0.004	< 0.001	0.239	0.001 -	
0.08	< 0.001	0.004	< 0.001	0.221	0.001 -	
0.07	< 0.001	0.004	< 0.001	0.220	0.001 -	
0.06	< 0.001	0.004	< 0.001	0.208	0.001 -	
0.05	< 0.001	0.004	< 0.001	0.222	0.001 -	
0.04	< 0.001	0.004	< 0.001	0.227	0.001 -	

within each network to be faulty. To allow for comparison between sender and link faults, we choose the subset size of link faults to 4, 14, and 49 for the three MCCNs, respectively.

For each type of fault (sender or link) with a specific change magnitude Δp, we run SSVD-MSPC and LASSO-MSPC on the simulated packet data and record the run length and CPU time it takes each method to detect the change. We repeat this experiment for 200 times, which allows us to compute the average and standard deviation of run length and CPU time. The computer configuration is as follow: 4-core

Intel Core i7 2.2GHz CPU with Mac OS X platform. The results are summarized in Tables 4 and 5. Table 4 shows that SSVD-MSPC is highly scalable, i.e., the computational time only increases minimally from a small network of 5 nodes and 20 sender-receiver pairs to a large network of 50 nodes and 2450 sender-receiver pairs. Since standard deviation of the CPU time is very small and therefore it is not shown for clarify of presentation. Even for the large network, SSVD-MSPC is able to detect faults very quickly (~ 0.001 seconds). In contrast, LASSO-MSPC is significantly slower and non-scalable. In particular, LASSO-MSPC is unable to detect faults in the large MCCN within a reasonable timeframe: about 740 seconds (12 minutes) are needed to produce the test statistic for each monitoring window. This performance is expected because the computational complexity of LASSO-MSPC is $O(qm + q^3)$ as reported in the original paper, where q is the number of sender-receiver pairs in an MCCN, while the complexity of SSVD-MSPC is only $O(q^2)$.

Furthermore, Table 5 shows that SSVD-MSPC and LASSO-MSPC have similar ARL_1 performance in detecting sender/link faults for the small MCCN. For detecting sender faults in the medium MCCN, the performance of LASSO-MSPC is substantially worse. This is because a sender fault changes both the mean and covariance structure of packet data but LASSO-MSPC was design for detecting mean changes. For detecting link faults in the medium MCCN, the performance of LASSO-MSPC is comparable to SSVD-MSPC. This is because the sender-receiver pairs affected by the link fault do not share senders or receivers so that the covariance change of package data is not large enough to hurt LASSO-MSPC badly. Nevertheless, we can observe performance worsening of LASSO-MSPC as the fault magnitude becomes smaller. For fault detection in the large MCCN, LASSO-MSPC is incapable of returning results within a reasonable timeframe as previously discussed. Overall, SSVD-MSPC has

Table 5: Run length performance of fault/change detection: ARL_1(standard deviation)

(a) sender fault

Change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)	2.02(0.20)	1.00(0.00)	-
0.09	1.00(0.00)	1.00(0.00)	1.00(0.00)	4.20(0.44)	1.00(0.00)	-
0.08	1.00(0.00)	1.00(0.00)	1.00(0.00)	8.48(1.06)	1.00(0.00)	-
0.07	1.00(0.00)	1.02(0.02)	1.00(0.00)	27.60(3.21)	1.00(0.00)	-
0.06	1.04(0.03)	1.00(0.00)	1.02(0.02)	105.48(14.67)	1.04(0.03)	-
0.05	1.08(0.05)	1.06(0.03)	1.04(0.02)	130.46(18.61)	1.24(0.08)	-
0.04	1.20(0.09)	1.20(0.08)	1.22(0.09)	131.36(18.49)	1.62(0.15)	-

(b) link fault

Change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.09	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.08	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.07	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.12(0.05)	1.00(0.00)	-
0.06	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.18(0.07)	1.00(0.00)	-
0.05	1.00(0.00)	1.02(0.02)	1.08(0.05)	1.60(0.15)	1.22(0.08)	-
0.04	1.24(0.07)	1.14(0.05)	1.42(0.10)	2.84(0.30)	3.44(0.33)	-

excellent ARL_1 performance across both sender and link faults and across different fault magnitudes.

3.4.2 Performance of SSVD-MSPC in fault identification

Following the monitoring phase is fault identification, i.e., once a change is detected and alarmed, packet data in the current monitoring window will be analyzed to identify the faulty sender-receiver pairs that are responsible for the change, which will further help trace to the faulty sender, receiver, or links. In this experiment, we assess the fault identification accuracy of SSVD-MSPC in comparison with LASSO-MSPC. We compute two accuracy metrics: sensitivity measures the proportion of faulty sender-receiver pairs correctly identified as such; specificity measures the proportion of non-faulty sender-receiver pairs correctly identified as such. Table 6 and 7 summarize the results. SSVD-MSPC achieves superior specificity across sender and link faults and across different fault magnitudes. The sensitivity of SSVD-MSPC is also universally high except for the link fault with the smallest magnitude (i.e., $\Delta p = 0.04$). In contrast, LASSO-SPC has difficulty balancing between sensitivity and specificity. For example, it achieves high sensitivity for the small MCCN at the price of low specificity, and achieves high specificity for the medium MCCN at the price of low sensitivity. This observation holds for both sender and link faults.

3.4.3 Performance of SSVD-MSPC in monitoring, fault identification and characterization for different temporal shapes of the fault

Following fault identification is fault characterization which estimates the temporal shape of the fault. Previous sections focused on the temporal shape that is a step change. In this section, extended experiments are conducted to include three typically occurring shapes in MCCNs, i.e., step, trend, and oscillating changes. Taking a sender fault as an example, a step change is a time-unvarying decrease of the in-control success probability of the sender, i.e., from p_0^S to $p_0^S - \Delta p$. Without loss of generality, we consider a linear trend change that decreases p_0^S to $p_0^S - atp_0^S$. An oscillating change decreases p_0^S to $p_0^S - r(t)p_0^S$, where $r(t)$ is a random variable. The magnitude of a

Table 6: Sender fault identification accuracy: mean (standard deviation)

Sensitivity

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.07(0.02)	1.00(0.00)	-
0.09	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.04(0.01)	1.00(0.00)	-
0.08	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.06(0.01)	1.00(0.00)	-
0.07	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.05(0.01)	1.00(0.00)	-
0.06	1.00(0.00)	1.00(0.01)	1.00(0.00)	0.03(0.01)	1.00(0.00)	-
0.05	0.95(0.02)	0.99(0.02)	0.97(0.01)	0.03(0.01)	0.99(0.00)	-
0.04	0.82(0.04)	0.97(0.03)	0.90(0.02)	0.02(0.01)	0.97(0.01)	-

Specificity

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	0.56(0.02)	1.00(0.00)	0.99(0.00)	1.00(0.00)	-
0.09	1.00(0.00)	0.56(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.08	1.00(0.00)	0.57(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.07	1.00(0.00)	0.54(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.06	1.00(0.00)	0.58(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.05	1.00(0.00)	0.59(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.04	1.00(0.00)	0.57(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-

step change is just Δp, while the magnitude of a trend or oscillating change is not as obvious. For the convenience of result presentation, we define the magnitude of a trend/oscillating change as the average change magnitude within the monitoring window. Two magnitudes are focused on in our experiment: 0.1 and 0.05 represent large and small changes, respectively. Also, we focus on sender faults in the medium MCCN, and report the performance of SSVD-MSPC in monitoring (ARL_1), fault identification (sensitivity and specificity), and fault characterization (cosine similarity between the true and estimated temporal shapes) in Tables 8, 9, 10 and Figure 13.

Table 7: Link fault identification accuracy: mean (standard deviation)

Sensitivity

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.37(0.03)	1.00(0.00)	-
0.09	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.36(0.03)	1.00(0.00)	-
0.08	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.26(0.02)	1.00(0.00)	-
0.07	0.98(0.01)	1.00(0.00)	0.99(0.00)	0.24(0.02)	0.99(0.00)	-
0.06	0.90(0.03)	1.00(0.00)	0.94(0.01)	0.19(0.02)	0.96(0.01)	-
0.05	0.72(0.03)	1.00(0.01)	0.83(0.02)	0.17(0.01)	0.87(0.01)	-
0.04	0.55(0.03)	0.93(0.03)	0.62(0.02)	0.16(0.01)	0.69(0.02)	-

Specificity

Fault/change magnitude Δp	Small MCCN		Medium MCCN		Large MCCN	
	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC	SSVD-MSPC	LASSO-MSPC
0.1	1.00(0.00)	0.76(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.09	1.00(0.00)	0.75(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.08	1.00(0.00)	0.74(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.07	1.00(0.00)	0.75(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.06	1.00(0.00)	0.76(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.05	1.00(0.00)	0.72(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-
0.04	1.00(0.00)	0.77(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	-

LASSO-MSPC is not performed in this experiment because it was not designed for detecting and characterizing trend and oscillating faults. The results show excellent performance of SSVD-MSPC in monitoring, fault identification and characterization across different fault magnitudes and different temporal shapes of the faults.

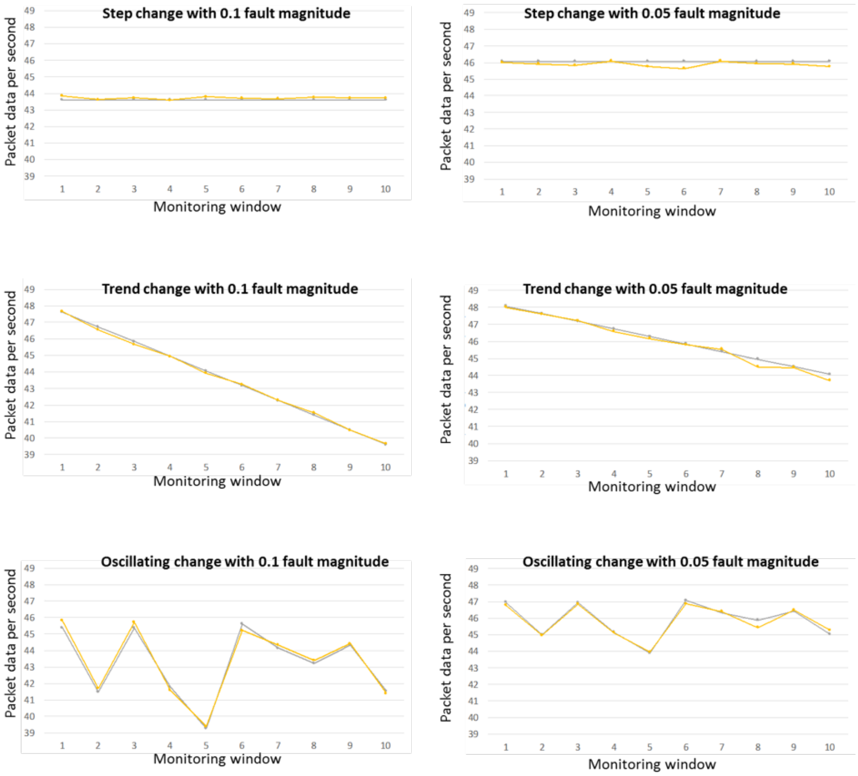


Figure 13: Comparison between true (black) and estimate (yellow) temporal shapes of the faults

Table 8: Run length performance of SSVD-MSPC: ARL_1 (standard deviation), for three temporal shapes of faults

Fault/change magnitude	Step change	Trend change	Oscillating change
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)
0.05	1.04(0.02)	1.02(0.02)	1.00(0.00)

Table 9: Fault identification accuracy for three temporal shapes of faults

Sensitivity

Fault/change magnitude	Step change	Trend change	Oscillating change
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)
0.05	0.97(0.01)	0.99(0.00)	1.00(0.02)

Specificity

Fault/change magnitude	Step change	Trend change	Oscillating change
0.1	1.00(0.00)	1.00(0.00)	1.00(0.00)
0.05	1.00(0.00)	1.00(0.00)	1.00(0.00)

Table 10: Cosine similarity between estimated and true temporal shapes

Fault/change magnitude	Step change	Trend change	Oscillating change
0.1	1.00	1.00	1.00
0.05	1.00	1.00	1.00

3.5 Case Studies on QoC Metrics in Radiology

Applied on multiple data streams from MCCNs, the proposed SSVD-MSPC algorithm shows its advantages for monitoring, fault identification, and fault characterization. According to our comparisons among different multi-variate process control algorithms, the SSVD-MSPC has the widest capability and is the most efficient method. As mentioned in previous chapter, the DDD and RQD system produce multiple QoC metrics which help to measure the efficiency, timeliness behavior of the whole radiology department in Mayo Clinic. However, when looking at the reports, we noticed that, although the reports would provide a good visualization, it is difficult to directly find any potential issues from the metrics we have. The causes of such situation include:

1. Each metric may follow a unique distribution. For instance, the mean values

of patient waiting time and radiologist turnaround time are different. If we further break a metric by scanner, exam type or patient type, the diversity of distributions makes the manual monitoring extremely difficult.

2. The metrics have hierarchical structures. For instance, the overall turnaround time metrics also includes radiology turnaround time, patient waiting time, etc. The errors introduced by sub-metric may not cause any alert but would accumulate an error on a higher-level metric. It is essential to identify the temporal and spatial location of each alert. Such uniqueness on data streams is also out of the capacity of traditional MSPC methods.

The SSVD-MSPC algorithm is a non-parametric approach and could be applied on data streams with multiple distributions, and also, the SSVD-MSPC achieves high efficiency and ability in fault identification and in differentiation temporal shapes of the fault. Thus, the SSVD-MSPC fits the application of monitoring QoC metrics streams from DDD system in the radiology department.

Radiology at Mayo Clinic has years of experience in operating multiple types of exams with multiple scanners. To demonstrate the performance of SSVD-MSPC algorithm, we collected $166,498$ CT exam records from 2011 to 2017 with DDD systems. These exam records contains 118 types of CT procedures and were collected from 8 CT scanners on 2 sites. After removing records with missing metrics, duplicates, scanner maintenance records, and other types of records that are not suitable for monitoring, the total amount of pre-processed records reduces to $51,913$. The majority parts of the removed records are those with missing metrics and duplicates. However, we don't have any information on if the records are in-control or out-of-control. We consulted our collaborators at Mayo Clinic and they mentioned there was not any

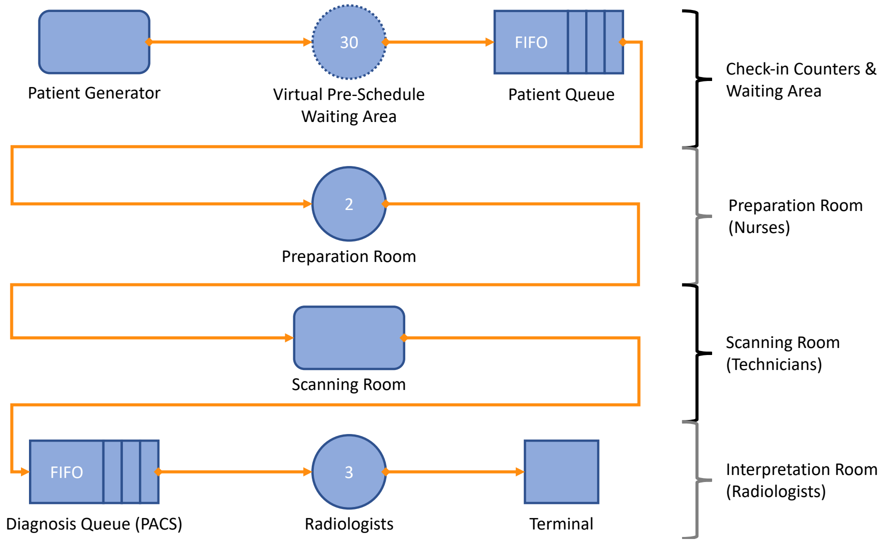


Figure 14: Radiologic Care Process Model

situation that had raised an alert during the past few years. Thus, we could assume that all the records we have are in-control.

3.5.1 Modeling the Radiologic Procedures

As illustrated briefly in Figure 1, the radiologic care process have multiple steps what would affect the metrics. We discussed the actual process that is used at Mayo Clinic, and further created the following model for the radiologic care procedure (see Figure 14).

A typical radiologic care process has multiple roles, for instance, a patient, a nurse, a technician, and a radiologist. When a patient comes to hospital, he / she will be checked in and waits in the waiting room until a nurse brings the patient into preparation room. In the preparation room, the patient will change clothes and the nurse will briefly go through the whole procedure with the patient, besides, the nurse will inject or ask the patient to take all necessary contrasts. At the same time, a technician will prepare the scanner for the coming exam. When scanner is ready, the patient will be taken to the scanning room and the technician will complete the scanning procedure followed by pre-process the images, and then clean the scanning room. After the exam take place, the patient will be checked out. The processed medical images will be stored in PACS. A radiologist will acquire the images for the patient and finish the interpretation.

The radiologic care process model (RCPM) is designed as a discrete event system to simplify the actual radiologic process. As shown in Figure 14, the RCPM starts with a patient generator which would determine how frequent a patient will come into the whole system and decide which type of exam the patient would take. From our analysis with DDD system, we have noticed that a patient come typically 30 minutes earlier than the scheduled time. We manually add a virtual step by a set of servers to compensate such patient behaviors. The patient will then enter the queue prior to the preparation rooms. When there is an available technician, the patient will enter the scanning room sub-system. In the scanning room sub-system, there are a number of parallel servers which are used to indicate scanners, and there also a number of technicians as "resources" which can be allocated to different scanning servers. The numbers of scanners and technicians can be different. After the scanning room, the patient will be checked out in reality, however, we reuse the entity as the scanned images for the corresponding patient and keep it in our model to finish the whole RCPM. We create another queue, named "diagnosis queue", as the PACS system, all scanned images sets are blocked and wait in the queue until a radiologist is available. After a radiologist serves the set of scanned images, the image set will leave the system.

3.5.2 Simulation of the RCPM

As we described previously, it is difficult for us to directly apply our algorithm on DDD data records since the statuses of in-control and out-of-control are unknown. We adopted the discrete events simulation approach in order to generate both data records of in-control and out-of-control with designed typical faults. To complete the simulation, multiple distributions are required:

1. We assume that inter-arrival time of patients follows an exponential distribution.
2. The simulation system assigns an exam type to each patient. The distributions of each exam types are also required.
3. We assume that the service times are all follows exponential distributions.

All mentioned distributions are estimated with all available DDD data records we have. However, due to the limitation of current HISs, there are also some of the service time distributions which are not available in DDD system. For instance, the preparation time of each exam, and the interpretation time of radiologist. We consulted our collaborator and use their estimations to complete those missing parameters in our simulation system. To simplify the simulation, we only generate the most common 10 types of CT exams. Table 11 summarizes the parameters and distributions we use for simulation.

We simulated 100 days of IC records, and for each day, the simulated time is 32400 seconds as nine-hour typical shift of radiologist. The records are summarized together, and 51 QoC metrics are derived from these IC records: 1) patient wait time, exam duration, technician turnaround time, diagnosis wait time (the time interval between image-ready and interpretation), radiologist turnaround time. These metrics

Table 11: Summary of Parameters in RCPM Simulation

Parameter	Location	Values
Inter-arrival Time	Patient Generator	$exp(737)$
Exam Type	Patient Generator	0.184 for “CHEST W-CST+d3D” 0.178 for “CHEST WO-CST+d3D” ...
Pre-schedule Time	Phantom Pre-schedule Waiting Area	$exp(1800)$
Number of Pre-schedule Server	Phantom Pre-schedule Waiting Area	Relatively large number (30)
Preparation Time	Preparation Room	$exp(3600)$ for “W-CST” $exp(1200)$ for “WO-CST”
Number of Preparation Server	Preparation Room	2
Scanning Time	Scanning Room	$exp(326)$ for “CHEST W-CST+d3D” $exp(214)$ for “CHEST WO-CST+d3D” ...
Post-scan Time	Scanning Room	$exp(621)$ for “CHEST W-CST+d3D” $exp(525)$ for “CHEST WO-CST+d3D” ...
Number of Scanner Server	Scanning Room	6
Number of Technician	Scanning Room	4
Interpretation Time	Radiologists	$exp(2593)$ for “W-CST” $exp(1758)$ for “WO-CST”
Number of Radiologist	Radiologists	3

are calculated daily for each of the top-10 exam types. 2) The daily patient volume. As we mentioned, these metrics are dependent among other metrics and we can find hierarchical structure between exam duration and technician turnaround time metrics. If some of the exam types do not exist on a day, the metrics would not be derivable and are replaced with global mean value of the metrics.

3.5.3 Performance of SSVD-MSPC on Radiology QoC Metric Data Streams

To demonstrate the performance of SSVD-MSPC algorithm, we apply the proposed method on data stream window in time range of a week, that is, the input of the algorithm would be a 51 metrics × 7 days matrix. We also assume that metrics a temporal independent without any embedded seasonality. Such assumption is valid since the research on ARIMA model and temporal seasonality are quite mature that we can always pre-process the given data streams to eliminate the seasonality and dependence. The IC control limit is estimated with similar approach on MCCN experiments, $\alpha = 0.02$ is used, which yields an in-control ARL of 50, and $UCL_{s_1}(0.02) = 19.77$.

3.5.3.1 Out-of-Control Fault 1: Insufficient Radiologists

Insufficient radiologist is a situation that there are not enough radiologists who are on their shift. Such situation may happen when several radiologists are on vocation or shift schedule is not optimal. For such out-of-control data, we manually reduced the number of radiologists from 3 to 1. Apparently, it would increase the "diagnosis wait time" and eventually decrease the daily patient volume. Table 12 shows the performance results from 1000 runs of SSVD-MSPC on "insufficient radiologists" fault.

Table 12: Performance on “Insufficient Radiologists” Fault

ARL_1	Overall Accuracy	Sensitivity	Specificity
1(0)	0.817(0.001)	0.948(0.002)	0.781(0.001)

As indicated, SSVD-MSPC algorithm archived perfect ARL_1, and the sensitivity of faulty metrics identification is 0.948 with 0.002 of standard deviation, and the specificity is 0.781 with 0.001 of standard deviation. The false-negative identifications happen when volumes of specific type of exams are extremely low. For instance, if “CHEST W-CST+d3D” exam does not exist on a certain day, the metrics of such exam will be replaced with global mean value which are in-control.

3.5.3.2 Out-of-Control Fault 2: Radiology Trainee

Radiology trainee will come to Mayo Clinic for training each year for core training and sub-specialty training. The trainees are less experienced and expected spend longer time to finish radiologic procedures, such as interpretation. According to (DeSimone *et al.*, 2017), the fluoroscopy time for radiology trainee is about 1.3 times long than it for faculty radiologists. We assume that for interpretation, the ratio of time follows the similar trend. In the simulation system, we changed the mean value of interpretation time $t_{interpretation}$ to $1.3 \times t_{interpretation}$. We expected a lower daily volume of patients, but longer interval of “diagnosis wait time” as well as “radiologist turnaround time”. Table 13 shows the performance results from 1000 runs of SSVD-MSPC on “trainee radiologists” fault.

As indicated, SSVD-MSPC algorithm archived 3.515 of ARL_1, and the sensitivity

Table 13: Performance on “Trainee Radiologists” Fault

ARL_1	Overall Accuracy	Sensitivity	Specificity
3.515(0.098)	0.865(0.002)	0.987(0.001)	0.692(0.004)

Table 14: Performance on “Insufficient Technicians” Fault

ARL_1	Overall Accuracy	Sensitivity	Specificity
49.534(1.503)	0.456(0.004)	0.727(0.005)	0.068(0.004)

of faulty metrics identification is 0.987 with 0.001 of standard deviation, and the specificity is 0.692 with 0.004 of standard deviation.

3.5.3.3 Out-of-Control Fault 3: Insufficient Technicians

Similar to the first type of OoC fault, insufficient technician may also occur. For such out-of-control data, we manually reduced the number of technicians from 4 to 2. Affected metrics includes “patient wait time”, “diagnosis wait time”, as well as “patient wait time”. Table 14 shows the performance results from 1000 runs of SSVD-MSPC on “insufficient technicians” fault.

As indicated, the performance of SSVD-MSPC algorithm is low, with 49.5 of ARL_1 and the overall accuracy of fault metric identification dropped to 0.456 with 0.727 of sensitivity and 0.068 of specificity. The ARL_1 is quite closed to ARL_0, meaning that the whole system under “insufficient technician” simulation produced very closed to IC system. After further analysis on the available technician in system, we noticed that, although we deployed 6 scanning room, 2 technicians would be also sufficient for the systems since the technician turnaround time is shorter compared to pre-scan time interval, that is, when a patient finished preparation, there would be an available technician for scanning. Under current setting, the SSVD-MSPC cannot distinguish a system with 2 technicians (designed OoC system) from a system with 4 technicians (designed IC system). However, with the performance results, the algorithm led us to further analysis the cause and found out an optimal number of technicians.

3.6 Conclusion

We proposed SSVD-MSPC for monitoring, fault identification, and fault characterization of high-dimensional packet data in MCCNs. These capabilities are critically important for QoS assurance of MCCNs whose malfunction or failure will result in serious social, economic, and/or security impacts. We showed that SSVD-MSPC is highly efficient and scalable in monitoring high-dimensional high-throughput packet data and has excellent ARL performance for detecting faults of different magnitudes in small, medium, and large networks. SSVD-MSPC also achieved high accuracy in fault identification and in differentiation of different temporal shapes of the fault.

We also applied the proposed SSVD-MSPC algorithm for monitoring and fault identification on radiology QoC metric data streams collected from simulation system. The SSVD-MSPC is an efficient method on monitoring QoC metrics and achieved good accuracy. However, we also noticed that with current simulation systems and designed faults, the specificities are relatively low. The reasons include: 1) the complex and hierarchical structure of data streams makes it more difficult to identify temporal locations; 2) the fault shift magnitudes on QoC data streams are vary, such difference in changes limits the capacity of LASSO procedure in SSVD-MSPC.

Future work may include applying SSVD-MSPC to other domains that generate high-dimensional data streams with complicated data distributions and correlation structures and that efficient online monitoring and fault detection is much needed. Considering the limitation of SSVD-MSPC algorithm shown on radiology QoC data, a better procedure for fault identification should be considered for different magnitudes of changes.

Chapter 4

MULTI-MODALITY DEEP TRANSFER LEARNING FOR COMPUTER-AIDED DIAGNOSIS

4.1 Introduction

Computer systems become one of the most important tools in healthcare, they can be used to store and visualize patients' information and records, to archive scanned images, to facilitate the ordering and scheduling, to analysis existing data, and to provide assistance of medical image interpretation and diagnosis (Wang *et al.*, 2017; Doi, 2007; van Ginneken *et al.*, 2011; Joo *et al.*, 2004; Giger *et al.*, 2013). Growing rapidly, computer-aided diagnosis is under the spotlight in both clinical and academic research, especially is applied to differentiate the malignancy and benignancy for lesions and tumors (Armato and Sensakovic, 2004; Sun *et al.*, 2013; Schwedt *et al.*, 2015; Hu *et al.*, 2015).

The conventional machine learning (ML) technique has been introduced to lesion/tumor detection area for decades. A typical design of conventional classification framework requires three major steps: feature extraction, feature selection and classification (Cheng *et al.*, 2016). Each step need to be addressed separately with proper technique to provide optimal results. The extracted features of medical images include but not limited to first order statistics (average, standard deviation, etc.), morphological metrics (roundness, size, volume, etc.), and texture features extracted with image processing algorithms (gray level co-occurrence matrix (GLCM)(Haralick *et al.*, 1973), local binary pattern (LBP)(Ojala *et al.*, 1994), etc.). More features can help algorithms represent medical images more accurate but the "curse of dimensionality" can also lead the algorithm to failure. Feature selection techniques, also known as dimension reduction methods, such as principle component analysis (PCA)(Pearson, 1901) and least absolute shrinkage and selection operator (LASSO)(Tibshirani, 1996) , are designed to trade off accuracy with smaller number of features so that ML algorithms (random forest, K-mean classifier, support vector machines (SVM), etc.) can be applied on selected features efficiently and get good performance. However, on the other hand, these three steps are not independent in the whole framework. Each technique used in ML framework has its own advantages but the overall best technique does not exist. Different combinations of techniques have to be resolved in order to find an optimal approach with best performance for the specific problem. Such procedure is complicated and exhausting.

The deep learning (DL) techniques are not new but become one of the fastest growing field again since the computing power grows. The fundamental units in DL networks are called perceptron, which was first introduced in 1957 (Rosenblatt, 1957) for binary classification problem. Mathematically, a perceptron can represent a function that maps the input $\mathbf{x}$ to and a binary output $f(\mathbf{x})$:

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

As shown in Equation 4.1, the final output of function $f(\mathbf{x})$ is based on the sign of a linear combination of element in vector $\mathbf{x}$ plus a bias number b. The binary output function is a composite function as shown in Equation 4.2.

$$f(\mathbf{x}) = h(g(\mathbf{x})), \quad \text{where } g(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \text{ and } h(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

The function of $g(\mathbf{x})$ is a regression function that mapping the input $\mathbf{x}$ to a real number z, the output z is also called a representation of $\mathbf{x}$, and the function of $h(z)$, also known as the *activation function*, is the binary classifier based on the sign of input value. Besides the binary activation function, there are several widely used alternatives, such as sigmoid, hyperbolic tangent, and rectified linear unit (ReLU). A typical example of DL networks is multi-layer perceptron (MLP) which stacks layers of perceptrons to mimic the behavior of neurons such as transmitting nerve impulses. In a single layer of MLP, there are several independent perceptrons that generate multiple outputs, such outputs are considered as the input vector of the followed layer. The formulation of a MLP can be modeled as follow:

Layer 1 (input layer):

$$o_{1,1} = h_{1,1}(g_{1,1}(\mathbf{x})), \quad o_{1,2} = h_{1,2}(g_{1,2}(\mathbf{x})), \quad \cdots, \quad o_{1,l1} = h_{1,l1}(g_{1,l1}(\mathbf{x}))$$

Layer 2 (hidden layer):

$$o_{2,1} = h_{2,1}(g_{2,1}(\mathbf{o}_1)), \quad o_{2,2} = h_{2,2}(g_{2,2}(\mathbf{o}_1)), \quad \cdots, \quad o_{2,l2} = h_{2,l2}(g_{2,l2}(\mathbf{o}_1))$$

Layer k (hidden layer):

$$o_{k,1} = h_{k,1}(g_{k,1}(\mathbf{o}_{k-1})), \quad o_{k,2} = h_{k,2}(g_{k,2}(\mathbf{o}_{k-1})), \quad \cdots, \quad o_{k,lk} = h_{k,lk}(g_{k,lk}(\mathbf{o}_{k-1}))$$

Layer m (output layer):

$$o_{m,1} = h_{m,1}(g_{m,1}(\mathbf{o}_{m-1})), o_{m,2} = h_{m,2}(g_{m,2}(\mathbf{o}_{m-1})), \cdots, o_{m,lm} = h_{m,lm}(g_{m,lm}(\mathbf{o}_{m-1}))$$

Figure 15 demonstrates the typical structure of an MLP model, each circle indicates a perceptron which is consist of a regression function and an activation function as we mentioned. The final output vector $\mathbf{o}$ is a new representation of the input vector $\mathbf{x}$. All perceptrons in the same layer accept the same input, but the weight vectors of

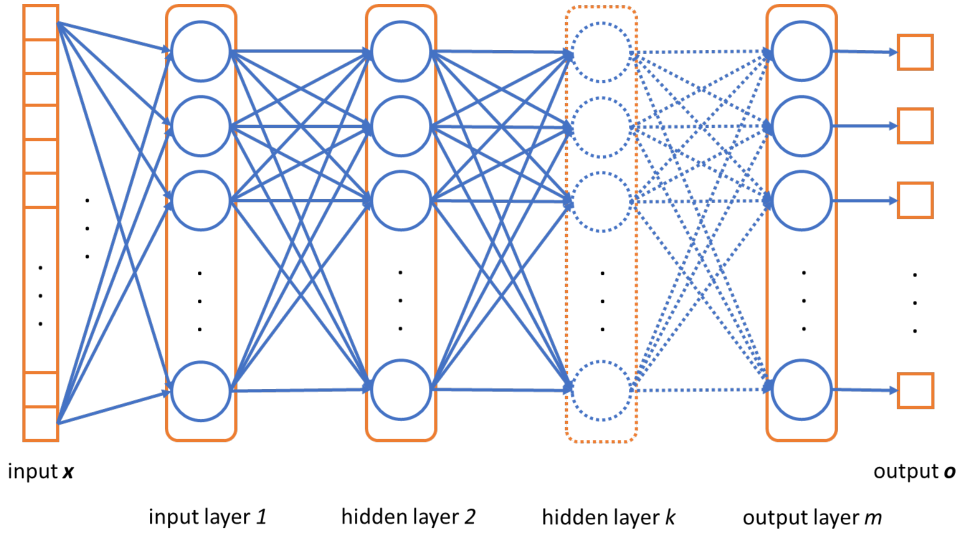


Figure 15: Typical Structure of MLP

each perceptron are independent. The function of MLP model is formed by composing many simpler regression functions.

The DL networks show their state-of-art power to archive better performance in image-related tasks such as image segmentations (Chen *et al.*, 2016; Shin *et al.*, 2013), and object recognitions (Szegedy *et al.*, 2016; He *et al.*, 2016). The medical imaging area adopted the DL techniques such as stacked auto-encoders (SAEs) and convolutional neuron networks (CNNs) for computer-aided diagnosis, mitosis detection, and organs detection. More details about DL frameworks in medical imaging area will be discussed in Section 4.2.

Breast cancer, due to its high incidence and high morbidity, contributes about 25% of death caused by cancers. In US, about 12% of women is and will be suffering from invasive breast cancer. The American Cancer Society recommends screening tests to women at average risk (those who doesn't have a personal history of breast cancer, a strong family history of breast cancer, or a genetic mutation known to increase risk of breast cancer, and has not had chest radiation therapy before the age of 30) (American Cancer Society, 2017). Full-field digital mammography (FFDM) is the most commonly used technique applied to breast cancer screening and diagnosis due to its low cost, high image quality and fast scan speed. However, FFDM suffers low sensitivity on plenty of cases, such as patients who have first-degree relative with a history of breast cancer (Kerlikowske *et al.*, 2000), have a known BRCA1 / BRCA2 gene mutation (Warner *et al.*, 2004), or are very young (Peer *et al.*, 1996). Such poor performance is difficult to improve since the difficulty to differentiate the suspicious breast lesions from dense and heterogeneous fibro-glandular tissues (FGT). Dynamic contrast-enhanced breast magnetic resonance imaging (DCE-MRI) is yet another solution for screening invasive breast cancer with much better performance (Berg *et al.*, 2012). However, compare to FFDM, the cost of DCE-MRI is much higher, and the scan time is also longer. These disadvantages limit the usage of DCE-MRI as the first choice of breast cancer screening. Contrast-enhanced digital mammography (CEDM) is a new imaging technique to maximize the advantages of both FFDM and DCE-MRI – having contrast-enhanced images to gain better diagnosis results and reducing the cost and scan time. A standard iodinated intravenous (IV) contrast agent is injected into breast and two series of scans with different X-ray energy levels are conducted during the CEDM imaging modality (Jong *et al.*, 2003). In general, CEDM generates images including LE images to provide similar information as FFDM images, and DES images as the substitute for DCE-MRI images. Diagnosis based on medical images heavily affects the effectiveness and patient safety of QoC in radiology department.

In this research, we are trying to analysis both types of CEDM images from 139 patients to differentiate the benignancy and malignancy, and assistant the radiologist to diagnosis without doing biopsy. For each patient, four types of images are considered – both LE and DES images from two scan angles (a top-down "cranial-caudal" (CC) view and an angled "mediolateral-oblique" (MLO) view). The images generated from CEDM imaging modality clearly have built-in coupling relationships between two series with different energy levels. Multi-modality learning is designed to fit classification models on a dataset in which multiple instances are considered as one single observation. For instance, images of the same patient should be considered as multiple instances of the same observation, and the same diagnosis result needs to be guaranteed on all these images. It is clearly that multi-modality learning technique is more appropriate than a learning technique which treats each image individually and independently.

We propose a novel deep-learning architecture with CNN to solve the multi-modality problems for breast cancer classification. Beyond taking advantages of deep learning architectures for state-of-the-art performance, there are multifold challenges we are facing:

- It is difficult to train a full deep CNN model from scratch since our data size is limited. As a well-addressed research topic, the ImageNet visual recognition challenge (Russakovsky *et al.*, 2015) provides over 1.4 million labeled images of real-world objects belonging to 1000 categories. Compare to that, our dataset of mammography contains only 139 patients. The possibility of "borrowing" the knowledge from large dataset result is an essential issue for applying CNN architecture.
- Each patient has 4 different images from two modalities of two angles respectively. How we can leverage the shared information among images but also keep uniqueness is another challenge.

- Regions of interest (ROIs) have been used on computer-aided diagnosis for years, and the accuracy of ROIs affects the final diagnosis results. However, defining ROIs requires experienced technologists and professional radiologists. If it is possible to limit the involvement of human resources and automate the end-to-end pipeline from scanned images to computer diagnosis is still unknown.

4.2 Literature Review

The deep learning techniques were introduced to medical image analysis in recent years and have shown promising results on different applications, such as image segmentation (Shin *et al.*, 2013; Liao *et al.*, 2013) and image classification (Xu *et al.*, 2014; Suk *et al.*, 2014). Focusing on image classification, researchers utilized different models from scratch to solve the computer-aided diagnosis problems. One of the strategy is using deep learning models as feature extractors, and then applying specific machine learning algorithm on the newly learned representation in feature space for classification. (Suk *et al.*, 2014) used Deep Boltzmann Machine (DBM) to find a latent hierarchical feature representation from MRI and positron emission tomography (PET) patches and then applied SVM models for the diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI). (Liu *et al.*, 2014) designed a deep learning architecture contains stacked auto-encoders (SAEs) to learn a representation of MRI images and aid the diagnosis of AD and MCI with multi-class SVM. (Cheng *et al.*, 2016) analyzed the ultrasound images to detect breast lesions and CT scans for pulmonary nodules by applying SAE.

Began from ImageNet Competition in 2012 (Russakovsky *et al.*, 2015), a new line of research which applied convolutional neural networks (CNNs) on image classification appeared and quickly became the mainstream for solving computer vision tasks, such as general classification. The performance of CNN classifiers was significantly improved by deploying deeper architectures (Simonyan and Zisserman, 2014; Szegedy *et al.*, 2015) since 2014. In deep CNN models, multiple convolution kernels are stacked as computing layers and contained millions of parameters that need to be estimated during the training process. Such complex models require a large amount of training data so that the parameters would not be ill-estimated. Take ImageNet competition for example, the models of general classification task of 1000 categories are trained with 1.4 million of open domain natural images. However, in the field of medical images, acquiring image data are costly and sometime impossible since the rareness of certain diseases.

Another widely used strategy for medical image classification adopts transfer learning techniques. Transfer learning is a technique which applying stored knowledge learned from one task to another similar one. For instance, human can learn how to distinguish one type of objects (cats) from other types of objects (dogs, fishes, etc.). When a radiologist trying to differentiate benign tumors and malignant tumors, the experience (knowledge) of distinguishing objects would help. Deep transfer learning architecture is commonly designed with pre-trained deep neural networks on a large dataset and a specific smaller dataset of interest for fine-tuning (continued training). Apparently, in the context of image classification, the ImageNet models are excellent deep networks to start with. Several researches worked on transferring knowledge from ImageNet classifiers to the computer-aided diagnosis tasks with different types of medical images. (Zeng and Ji, 2015) combines a pre-trained VGG model (Simonyan and Zisserman, 2014) and then added 3 convolutional layers and 2 fully connected layers to solve a multi-task classification problem with in situ hybridization (ISH) images of mouse brains. Similar approach with VGG model and 4 following layers model was applied to ISH drosophila images in (Zhang *et al.*, 2016).

Mammography images are wildly used to screen breast cancer. To the best of our knowledge, computer-aided mammography screening in the deep learning context is still new to the society. Dhungel *et al.* combined R-CNN (Regions with CNN features) technique and random forests classifier to detect and classify breast masses in standard FFDM images (Dhungel *et al.*, 2015a,b). Based on manually segmented ROIs on standard FFDM images and several steps of preprocessing, Arevalo et al. proposed a CNN classifier with 4 layers of convolution feature maps for mass lesion classification (Arevalo *et al.*, 2016). Lévy and Jain utilized pre-trained CNN models with fine-tuning on mammography data and shown that the knowledge can be effectively transferred on mass classification tasks (Lévy and Jain, 2016). The datasets for these researches include images taken from MLO and/or CC views, however, the multi-instance information and relationship were not accounted in the final model.

In our research, the CEDM contains 4 mammography images from both MLO and CC views with LE and DES. The design of CEDM procedure raises the requirement for computer-aided diagnosis system to analyze multi-modality images simultaneously. Similar tasks of multi-modality classification with medical images have been discussed in several publications. Suk *et al.* combined both MRI and PET images and applied them together on SAE (Suk and Shen, 2013) and DBM (Suk *et al.*, 2014) models to identify AD / MCI cases. (Zeng and Ji, 2015) used multiple sagittal sections from mouse brain ISH images and adopted a share-weight tuned VGG model to get the concatenated multi-modality representation for the proposed multi-task classifier. (Xu *et al.*, 2014) proposed a deep learning framework for patch-level classification of high resolution histopathology images and shown that the multi-modality algorithm is effective and efficient as a weakly supervised classifier.

As seen, several researches have been done and using different deep learning architectures to learn specific representations of medical images to directly used as end-to-end computer-aided diagnosis classifier. Some of these afore-reviewed works applied transfer learning techniques solving classification tasks on other diseases / tumors. However, for mammography images, these researches are not directly applicable because of the following reasons: 1) The multi-modality tasks haven't been well addressed in mammography. The image pairs of the same scan view share location information of the breast mass and the image pairs of the same energy level may contain relatively "close" representation, but such information were ignored. 2) The existing works are applied on standard FFDM images which is not an optimal tool for breast cancer screening. The multi-modality CEDM images have unique information but we have not find many related works. 3) Most of the aforementioned researches used manually defined ROIs and the process of drawing ROIs is both time and labor consuming. We are looking for finding a way that could use less human involvement but be comparably efficient.

4.3 Proposed Method

As mentioned before, most of the existing deep learning networks are designed to solve the open-domain image classification on natural images. However, the medical images have their uniqueness. Compare to natural images, the medical images also have shapes, brightness, as well as recognizable objects. However, the objects in medical images are captured with special designed devices with invisible lights or special waves, such as X-ray, gamma ray, and ultrasounds. Beyond images with visible lights, the medical images also contain unique information such as density (X-ray), activity levels (functional MRI) and protein levels, such information is embedded in color or brightness for human sights. To account such features in medical images, we proposed a transfer-learning architecture to utilize the medical images features without losing knowledge learned from natural images. More specific, we started with deep CNN networks, called InceptionV3 models, trained with ImageNet dataset, and trimmed several uppermost layers to get top-less pre-trained models. Then, we added our own convolution layers on top of the pre-trained models and constructed a two-class classifier. The whole model is trained with CEDM images to learn parameters in the classifier and tune parameters in the pre-trained models.

4.3.1 Pre-trained Model for Natural Images

The InceptionV3 model is redesigned from previous version (Inception Model, also known as GoogLeNet (Szegedy *et al.*, 2015)) by substituting the 5×5 convolutional kernel with two stacked 3×3 convolutional kernels to reduce the computational complexity and the number of parameters. The new Inception module is illustrated in Figure 16.

In each Inception module, the convolutions shown with dashed boarder are designed as low-dimensional embeddings. However, these embeddings are dense, compressed form. Such representation is difficult to model (Szegedy *et al.*, 2015). Thus, the 1×1 convolutions (shown with dash-dot boarder) are added to the module to keep the representation sparse and be used as rectified linear activation functions. Additionally, a pooling component is also included as a parallel path since the operation is currently

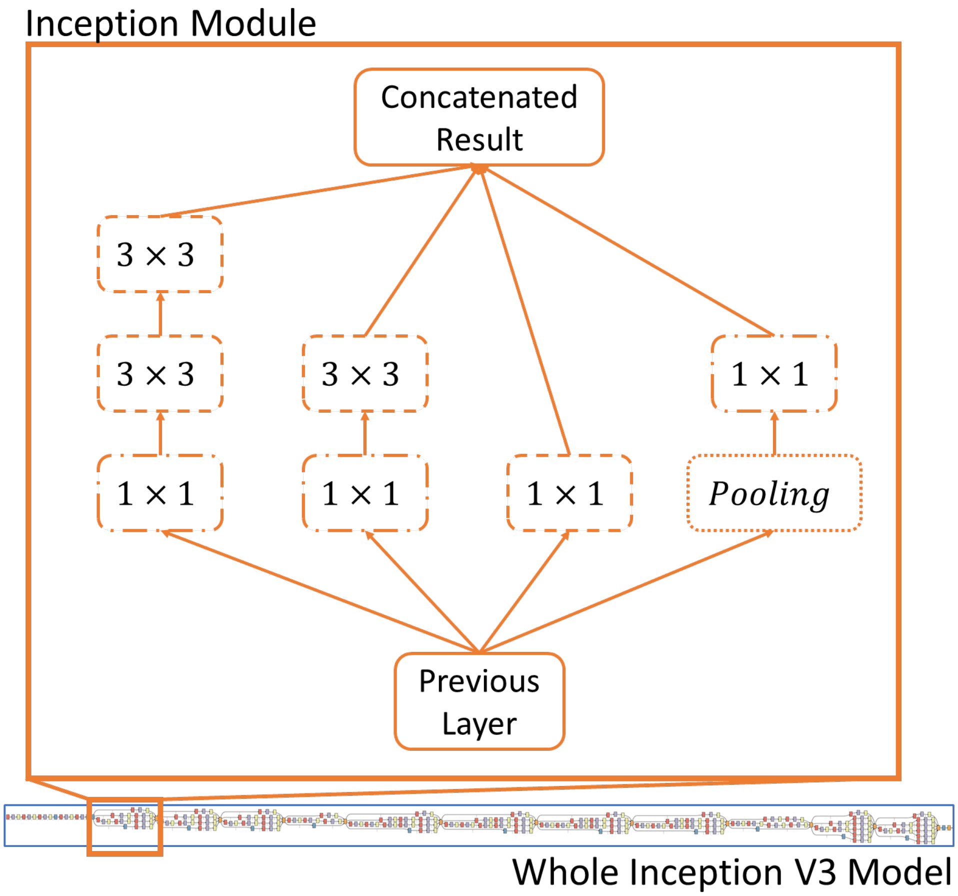


Figure 16: Structure of Inception Module in InceptionV3 Model

essential for success. The complete Inception model stacks 10 modules and put a simple classifier with a global average pooling layer followed by a full connected dense layer with a $1000D$ output vector.

In our proposed model, we truncate the top full connected layer from the complete Inception model since we are trying to solve different classification task. Thus, the output of Inception model is a $2048D$ vector with embedded information of the original input images.

4.3.2 Calibrated Multi-Modality Deep Learning Architecture for CEDM Images

A key challenge in applying CNNs to medical images is that it is difficult to collect enough data observations especially those with labels. One reason causes the lack of training data is the cost of devices and acquiring images. The breast MRI machine including coils is approximate $\$685,000$ and the cost of CEDM is approximately $\$370,000$ (Patel *et al.*, 2017). Another reason is the rareness of diseases. For instance, the breast angiosarcomas account about 0.04% of all breast cancers, or approximately one in every 2500 patients (Kaklamanos *et al.*, 2011). To overcome the limitation caused by data shortage and develop a promising architecture for medical images, transfer learning techniques is our recommendation to preserve the knowledge that have been learned from natural images and will be calibrated with new knowledge in CEDM images. The transfer learning techniques have yields good performances on several classification tasks with natural images (Zeiler and Fergus, 2014)(Oquab *et al.*, 2014).

Different from standard FFDM, the CEDM images are taken with two levels of radioactive energy – low energy similar to FFDM and high energy. Each CEDM procedure generates 4 images from 2 angles with 2 different energy levels for each patient. Figure 17 shows a sample set of CEDM images for patient “cedm51m18”. As illustrated, the DES images on the left-hand side enhanced the contrast between normal tissues (gray background) and potential tumor masses (brighter areas). The images clearly indicate the location and shape of the mass. In contrast, the images on the right-hand side capture more texture details of vessels, fibro-glandular and skins, however, the tumor mass are hidden in textures and difficult to differentiate from

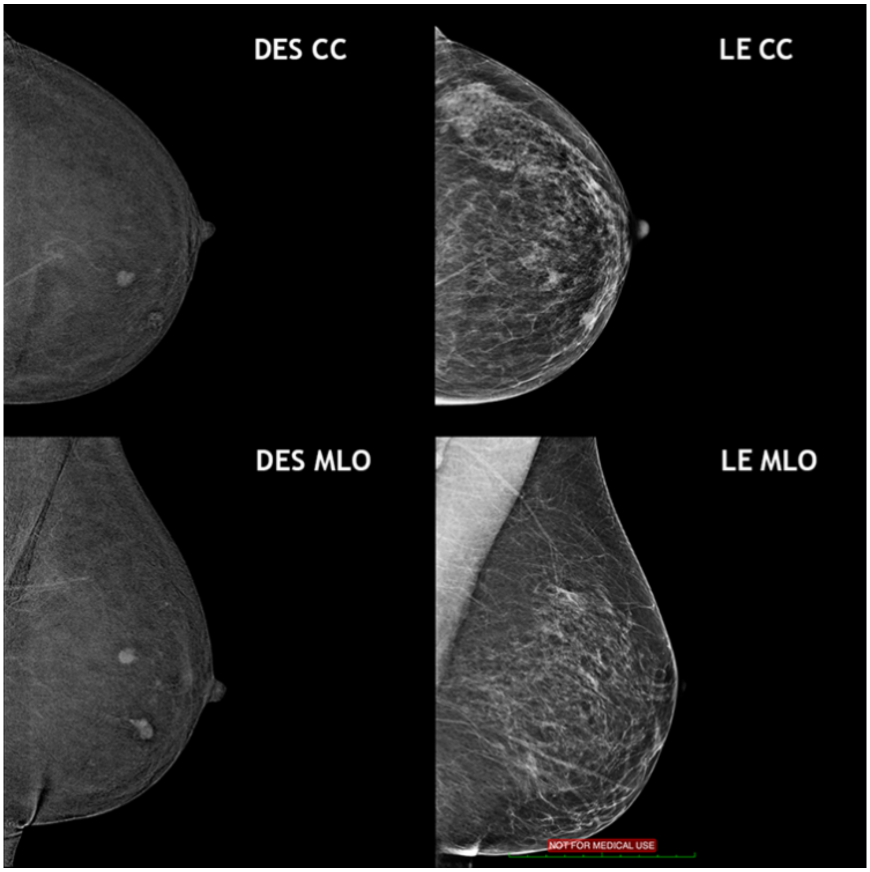


Figure 17: Sample Images of CEDM with Dual Angles and Dual Energy Levels

other tissues. Clearly, different energy levels produce two types of medical images which have unique perceived features.

A simple way to learn image representation is applying a pre-trained Inception model as a feature extractor. The tuning procedure during the learning process will update parameters of the pre-trained model and calibrate the model to fit images of CEDM. However, such simple model has limitation of differentiate representation features between two different CEDM modalities. In order to learn a powerful representation model for CEDM images, we proposed to deploy two separated InceptionV3 models, one for DES image pair and the other for LE image pair. The dual model

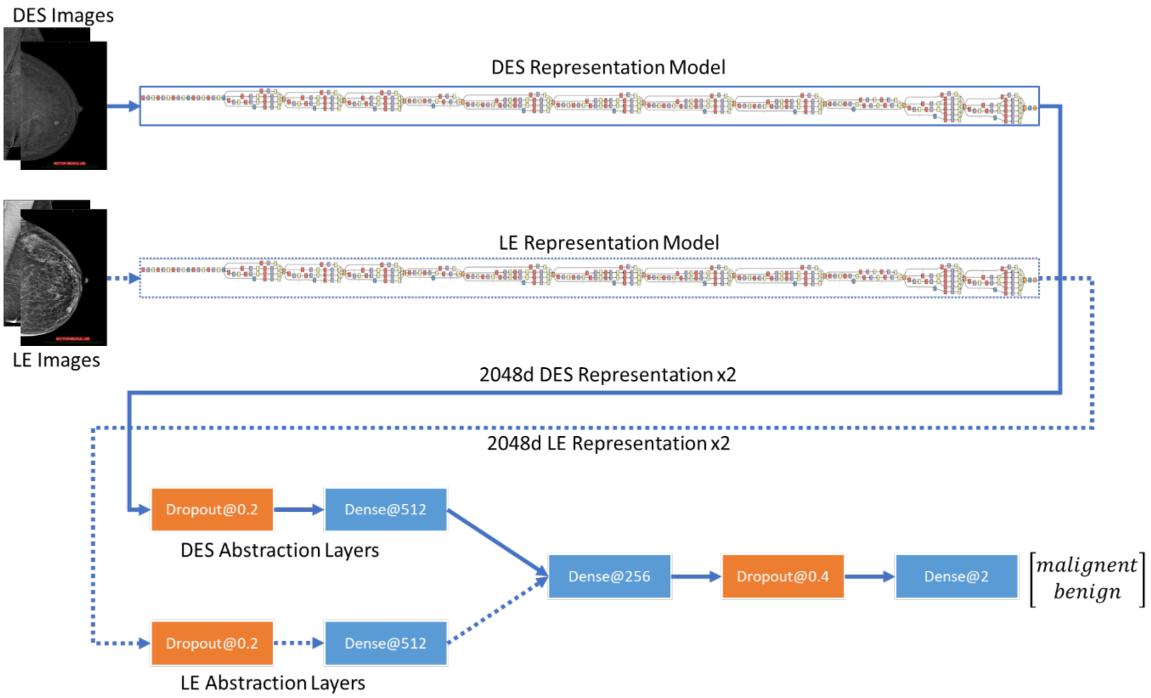


Figure 18: Proposed Calibrated Multi-Modality Model for CEDM Images

framework is expected to handle the perceived features of each modal of images independently.

As illustrated in Figure 18, our proposed multi-modality model consists with two major part. The first part contains two separate pre-trained InceptionV3 models designed for DES and LE images respectively. The input of pre-trained model part is original CEDM images or patches and the output is $2048D$ representation vectors. Take DES image path for example, since DES modality contains two images taken from different angles, the DES representation model will get two vectors. These two vectors are concatenated as a single $4096D$ vector and then sent to DES abstraction layers with a dropout layer and a full-connected dense layer with 512 output nodes. That is, a pair of DES images will be represented as a $512D$ vector after the first part of model. The same procedure is also applied to LE image path. The second part of model has 3 layers: full-connected layer reduced the dimensionality from $1024D$ (the concatenated vector of two outputs from abstraction layers) to $256D$; a dropout layer; and another full-connected layer to finally classify inputs into two classes – malignant and benign. The proposed model has $48,063,830$ parameters, and each InceptionV3 model has $21,802,784$ parameters.

4.4 Experiments and Results

The radiology department of Mayo Clinic Arizona provides CEDM images in total from 139 female patients. Each image of the dataset has a corresponding manually drawn ROI indicating the region of breast tumor. After removing cases with insufficient images, wrong images, or wrong corresponding ROIs, we then selected all 48 benign patients and randomly selected 48 malignant patients to get a balanced dataset. This balanced dataset contains 96 cases and is randomly split into 3 sets: 60 patients as training samples, 16 patients as validation samples and 20 patients as testing samples. All subsets are also balanced. Since the sample size of our dataset is relatively small for deep learning task, we applied data augmentation techniques during the training and validation procedures. Details of dataset are listed in Table 15.

The manually-drawn ROIs (MD-ROIs) are in irregular shapes and have different size. Figure 19 indicates the histograms of the MD-ROI sizes in unit of pixels in original images (in total 278 tumors, 139 in CC view and 139 in MLO view). We can observe that the sizes of malignant tumor have a heavy tail on the right when MD-ROI size over $100,000$ pixels. The reason is that malignant tumors are more invasive and tend to be larger. The variant nature of ROIs makes it difficult to directly use the raw region for classification tasks. To utilize location information of the ROIs, we introduce

Table 15: Details of Datasets

Dataset	Number of Samples	Augmentation Parameters
Training	60×5	Shift ratio: $(-0.05, 0.05)$ Rotation: $(-3, 3)$ Zoom Ratio: $(-0.05, 0.05)$ Horizontal flip Vertical flip
Validation	16×5	Shift ratio: $(-0.05, 0.05)$ Rotation: $(-3, 3)$ Zoom Ratio: $(-0.05, 0.05)$ Horizontal flip Vertical flip
Testing	20	Original image

lazy-drawn ROIs (LD-ROIs) which do not require accuracy shape and size information of the region. In practice, the LD-ROIs used in our experiments are defined as patches with specific sizes that cropped around the center of manually-drawn ones. Besides, the usage of LD-ROIs also reduce the process time of professionals which makes them dual-purpose. Figure 20 shows the histogram of MD-ROI area percentage, which is the ratio between MD-ROI area inside of its corresponding LD-ROI and the patch size (256×256). As mentioned, the malignant tumors are larger than benign tumors, we can also notice that the percentage of tumor area in patch for malignant tumors are also bigger than benign cases. The majority percentage of MD-ROI in unisize patch of benign cases is $0\% \sim 30\%$, while it of malignant cases is $10\% \sim 30\%$ and $80\% \sim 100\%$.

All following experiments are performed on NVIDIA GeForce GTX 1080 GPU with 8GB graphic memory, the CPU is Intel Core i7 5930K. Each experiment consists of 10 runs, and each run has 100 epochs. During each epoch of training, the training

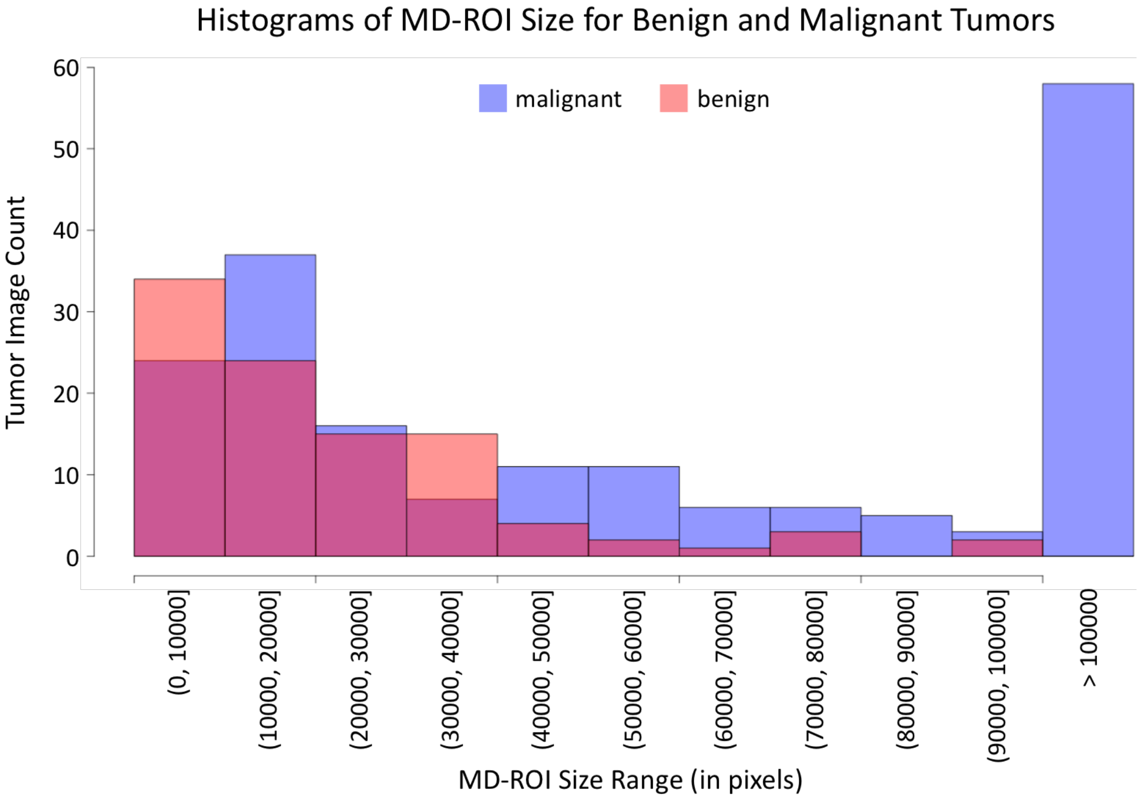


Figure 19: Histograms of MD-ROI Size for Benign (red) and Malignant (blue) Tumors

and validation sets are augmented to have 5 augmentations for each patient, and the order of patients is randomized. The testing procedure applied original images without augmentation.

In this section, we evaluate the performance of the proposed method and compare the effectiveness with several other architectures to analysis the efforts of 1) tuning the representation parameters as well as the classifier; 2) applying dual representation models instead of single representation models; 3) different patch sizes of lazy-drawn ROIs; and 4) shifted center location for lazy-drawn ROIs.

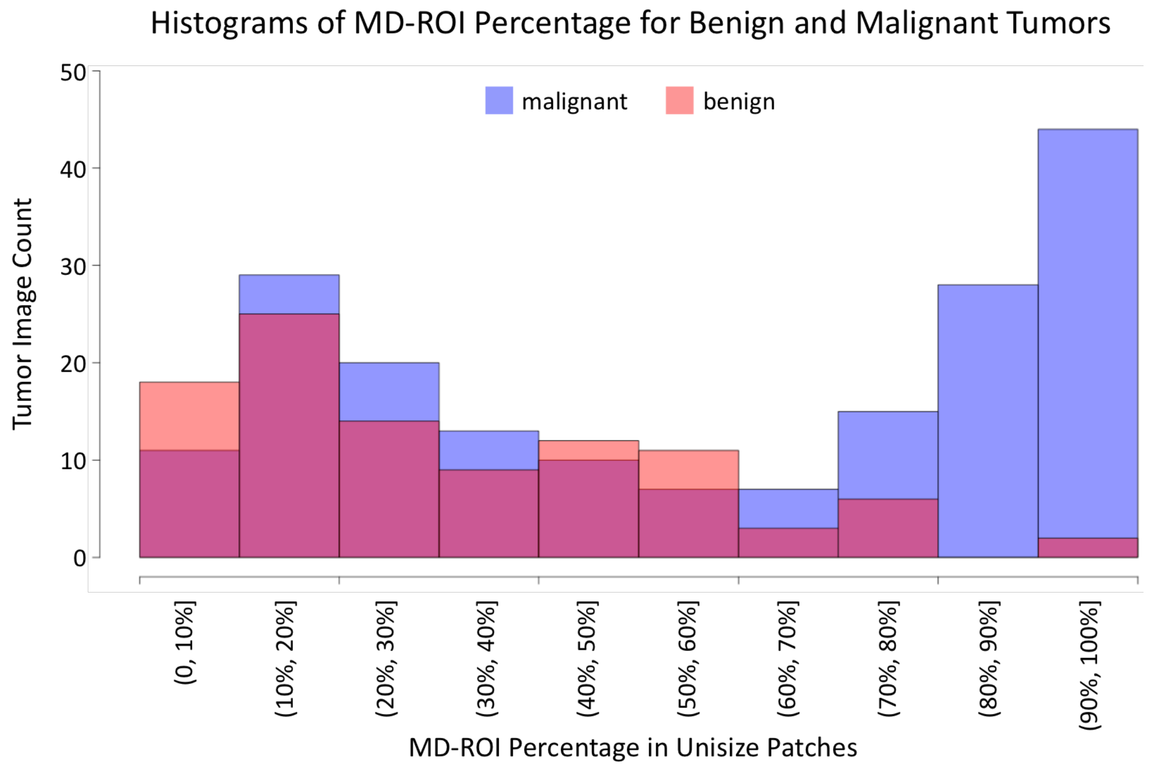


Figure 20: MD-ROI Histograms for Benign and Malignant Tumors

4.4.1 Tuning Parameters

In transfer learning architecture, we want to preserve as much as possible of the learned knowledge from the source task and at the same time learn as much as possible of the new knowledge from the target task. One strategy to build the transfer learning model is keeping the parameters of pre-trained but only estimating parameters in layers of classifier, we call it semi-tuning strategy. Another strategy is more aggressive by tuning all parameters in the designed model, named full-tuning strategy.

To compare the pros and cons of these two training strategies, we applied a simple model with only one InceptionV3 model for all images with both energy levels. More specific, the semi-tuning strategy on the simple model we used would update

Table 16: Comparison between Training Strategies on Simple Models

Tuning Strategy	Training Time	Training Accuracy	Validation Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
Semi-Tuning	$36s$	0.9800	0.5000	0.5000	1.0000	0.0000
Full-Tuning	$52s$	1.0000	0.7625	0.7000	0.6000	0.8000

$4,326,658$ parameters of the last 3 dense layers during training, but the full-tuning strategy would change $26,095,010$ parameters, which a about 5-time more parameters than semi-tuning model. It is obvious that training a deep transfer learning model using semi-tuning strategy require much less calculations than full-tuning model. The training speed is one of the advantages of semi-tuning strategy.

We independently trained 10 runs with each strategy, and in each run, 100 epochs were included. These models were trained with the same initialization (ImageNet pre-trained weights for InceptionV3 model and standard initial weights for other layers in classifiers), dropout rates, and learning rate policies, and they only differ in the random order of input images. We applied models on 256×256 crops of images and used a random split with serial number "M1GU0G". Table 16 lists the performance of models with different training strategies.

The final models were selected from top-10 training accuracy models with highest validation accuracy respectively. From our observation, top-10 models have very close performance in terms of training accuracy (~ 0.01 difference), however, the validation accuracy may vary. We believe such situation was caused by two reasons: 1) lack of training samples compromised the robustness; 2) random orders varied the final performance.

As we expected, the average training time of semi-tuning models is around

36s/epoch and of full-tuning models is around 52s/epoch. The semi-tuning model requires only about 70% of training time compared to full-tuning models. However, performance is an unavoidable shortage to semi-tuning models. Although the training accuracy reached 0.9800, the validation accuracy and testing accuracy are both only 0.5000. More specific, the testing sensitivity of malignant class is 1 while specificity is 0, which indicating that the classifier tends to assign all samples to malignant class. Such unbalanced performance is not acceptable on any tasks. The full-tuning strategy, by contrast, trained a model with perfect training accuracy and 0.7625 as validation accuracy. The testing accuracy is lower than validation, get a 0.7000. All these three accuracies metrics outperformed the semi-tuning models. Sensitivity and specificity of malignant class are relatively more balanced at 0.6 and 0.8, respectively. Compared these two training strategies, the full-tuning models should be selected to allow the CEDM images to adjust the weights / parameters in the pre-trained InceptionV3 model for better adaptation of medical images.

4.4.2 Single Representation Model and Dual Representation Model

In previous sub-section, we applied our dataset on a deep learning model with single InceptionV3 model for all 4 images of patients, that is, we used the same representation for all inputs. However, as in Figure 17, we showed that the DES images have different visual features with LE images. The CEDM technique is designed to use contrast-enhanced iodinated IV agent to provide better differentiation between tumor mass and normal tissues compared to FFDM. In our proposed architecture, we used dual-representation models to capture the uniqueness of images of each energy-level. The two representation models are tuned with certain type of images

Table 17: Performance Comparison between Single and Dual Representation Models

Rep. Strategy	Training Time	Training Accuracy	Validation Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
Single-Rep.	$52s$	1.0000	0.7625	0.7000	0.6000	0.8000
Dual-Rep.	$117s$	0.9933	0.7500	0.8000	0.8000	0.8000

respectively. Specifically, the DES representation model is tuned with only DES images of each sample while the LE representation model is tuned with LE images. In Table 17, we compared the performance of two models: model with single-representation and model with dual-representation.

As shown, the dual-representation model has slightly lower performance in term of training and validation accuracy. Since we have limited augmented dataset for training, the performance difference, if projected in number of samples, the differences are only 2 samples (out of 300) and 1 sample (out of 80) worse respectively. Comparing the metrics on the 20-sample testing set, the specificity, sensitivity and overall accuracy are all 0.8000 and two of these are better than single-representation model. According to this comparison, we conclude that the dual-representation model has its advantage to capture unique representation for both DES and LE images, and the slight gap on training and validation set can be compromised for a better performance.

4.4.3 Model Robustness

4.4.3.1 Patch Sizes of LD-ROIs

Another problem we want to address is what is the optimal patch size for LD-ROIs. We analyze and report the performance of multiple patch size choices, including 64 × 64, 128 × 128, 256 × 256, 512 × 512 and the original size. In real world, the sizes of tumors may vary. Figure 5 shows tumors from two different patients, manually-drawn ROIs for each tumor are included to help readers recognize the location and shape. We also put 4 rectangles in each of the image showing the crops we use as LD-ROIs. As illustrated, the red semi-transparent polynomial indicates the manually-drawn ROI tagged by radiologists, 4 green rectangles show the lazy-drawn ROIs sizing from 64 × 64 to 512 × 512, respectively. In the left LE image, the manually-drawn ROI is relatively small. The 64 × 64 and 128 × 128 LD-ROIs can cover mostly of the manually-drawn ROI, but the 256 × 256 and 512 × 512 patches also include partial of the normal tissues. Also notice that the original 512 × 512 exceeds the border of image, thus the lazy-drawn ROI for this case was shifted left to make sure the whole ROI is inside of the image. By contrast, if the manually-drawn ROI is relatively large, as shown in the right part of Figure 21, all 4 LD-ROIs focus on the mass itself and only a small part of normal tissues are included in 512 × 512.

Table 18 compares the performance of full-tuned dual-representation models on different sizes of lazy-drawn ROIs.

As listed, all models get high training accuracy over 0.99. The 256 × 256 LD-ROIs model has the highest validation accuracy as 0.7500 while the 128 × 128 LD-ROIs

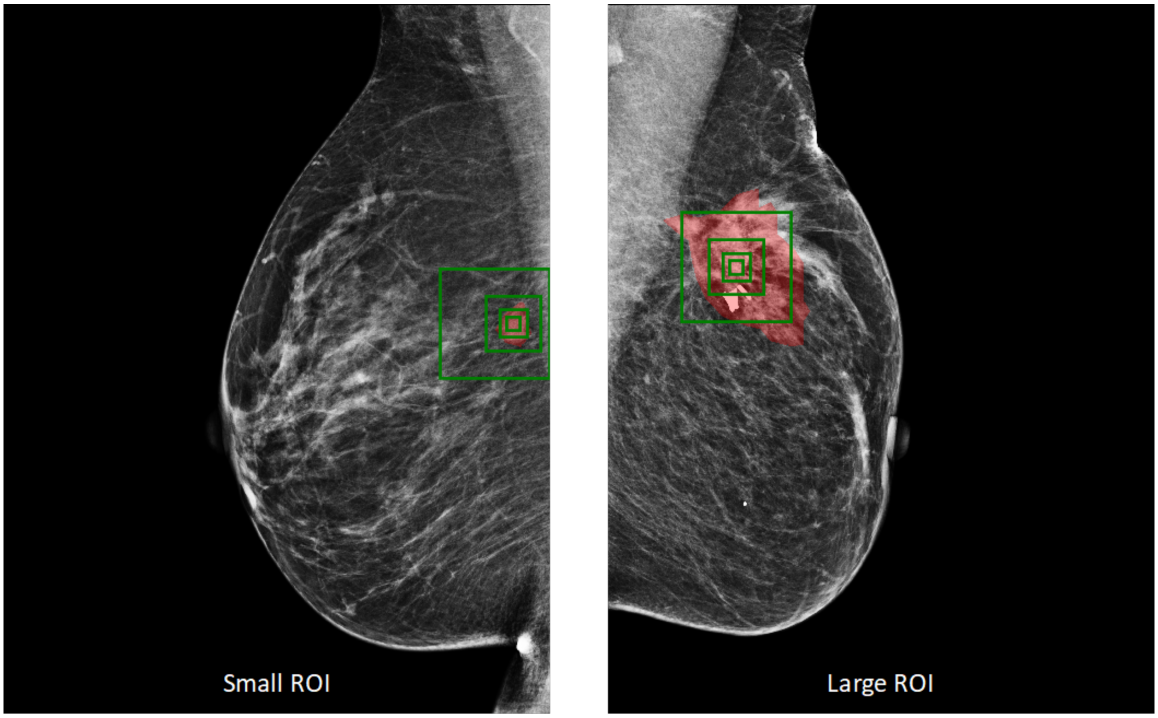


Figure 21: Comparison between Lazy-Drawn ROIs and Manual-Drawn ROIs

Table 18: Performance Comparison among Different Sizes of Lazy-Drawn ROIs

LD-ROI Size	Training Time	Training Accuracy	Validation Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
64 × 64	$29s$	0.9950	0.7063	0.7000	0.7000	0.7000
128 × 128	$33s$	0.9967	0.6813	0.6500	0.8000	0.5000
256 × 256	$117s$	0.9933	0.7500	0.8000	0.8000	0.8000
512 × 512	$228s$	0.9983	0.6938	0.7500	0.9000	0.6000

model has only 0.6813 of validation accuracy. In terms of the performance on testing set, the 256 × 256 model also has the best metrics with overall accuracy, sensitivity and specificity of all 0.8000. Based on the reported performance metrics, we conclude that the 256 × 256 LD-ROI size is optimal on our proposed dual-representation deep transfer learning architecture. For smaller tumors, small LD-ROI size may cover whole tumor area without including large proportion of normal tissues, however, if the tumor is relatively large, small LD-ROI size focuses only around the centroid of tumor mass and loss the shape and border information of tumors. On contrast, if the LD-ROI size is large, it would help capture all mass of large tumors, but for small tumor cases, there would be a large proportion of normal tissues. Both situation would lower the performance of final model. A mid-size LD-ROI is a compromised selection which won't mess tumor tissue with much normal tissue, which capture more information of tumors.

4.4.3.2 Shifted LD-ROIs

We further push our model to a dataset with shifted LD-ROIs to evaluate the robustness. The shifted LD-ROI is generated with the following mechanism: 1) load manually-drawn ROI information and get the center point c of ROI; 2) shift c to c' by increasing or decreasing 20 pixels on one of the axis; and 3) crop a LD-ROI by 256 × 256 square window. To maintain the coupling relationship between images of both energy levels, the shifts for the same angle of the same patient remain the same.

Table 19 compares the performance of our model on LD-ROI and shifted LD-ROI. As observed, on shifted dataset, the training and validation accuracies are slight better than original LD-ROI dataset. The overall testing accuracy remains the same at 0.8000

Table 19: Performance Comparison on Original LD-ROIs and Shifted LD-ROIs

ROI Center Position	Training Accuracy	Validation Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
Original	0.9933	0.7500	0.8000	0.8000	0.8000
Shifted by 20 pixel	0.9950	0.7938	0.8000	0.7000	0.9000

indicating robustness even with a 20-pixel shift ($\sim 7.8\%$ of the ROI size). However, we also addressed imbalance performance on two classes. The shift increased accuracy on benign class while compromising malignant class on the testing set. Figure 22 compares the histograms of tumor area in patch between original ROIs and shifted ROI. The histogram for original LD-ROIs is illustrated in red, while shifted LD-ROI is in blue, and the overlap of two histograms are in magenta. As we can see, the major part of the histograms are overlapped, and slight difference can be found when tumor size falls into $(70\%, 100\%]$ range. When tumors are relatively small ($< 50\%$), the shift of 20 pixel won't affect the percentage since most part of the patches are normal tissues; when tumors are extremely large (over 90%), the whole area of patch covers only partial of the whole tumor, thus the shift also won't change the percentage. When tumor sizes fall into $(50\%, 90\%]$, a 20-pixel shift may changes the percentage of MD-ROI. As illustrated, we observer that in $(70\%, 80\%]$, shifted LD-ROI has one more case than original LD-ROI; and in $(90\%, 100\%]$, shifted LD-ROI has one less case than original LD-ROI. Also should we notice that, since the histogram divided continues percentage into discrete bins, some changes may not reflected in the graph. Based on this comparison, we could conclude that the original LD-ROIs and shifted LD-ROIs follow the same distribution in term of the percentage of tumor. The Kolmogorov-Smirnov test returns D stat equals to 0.0035211, with $p = 1$. We rule out the distribution difference as one reason that cause the performance difference.

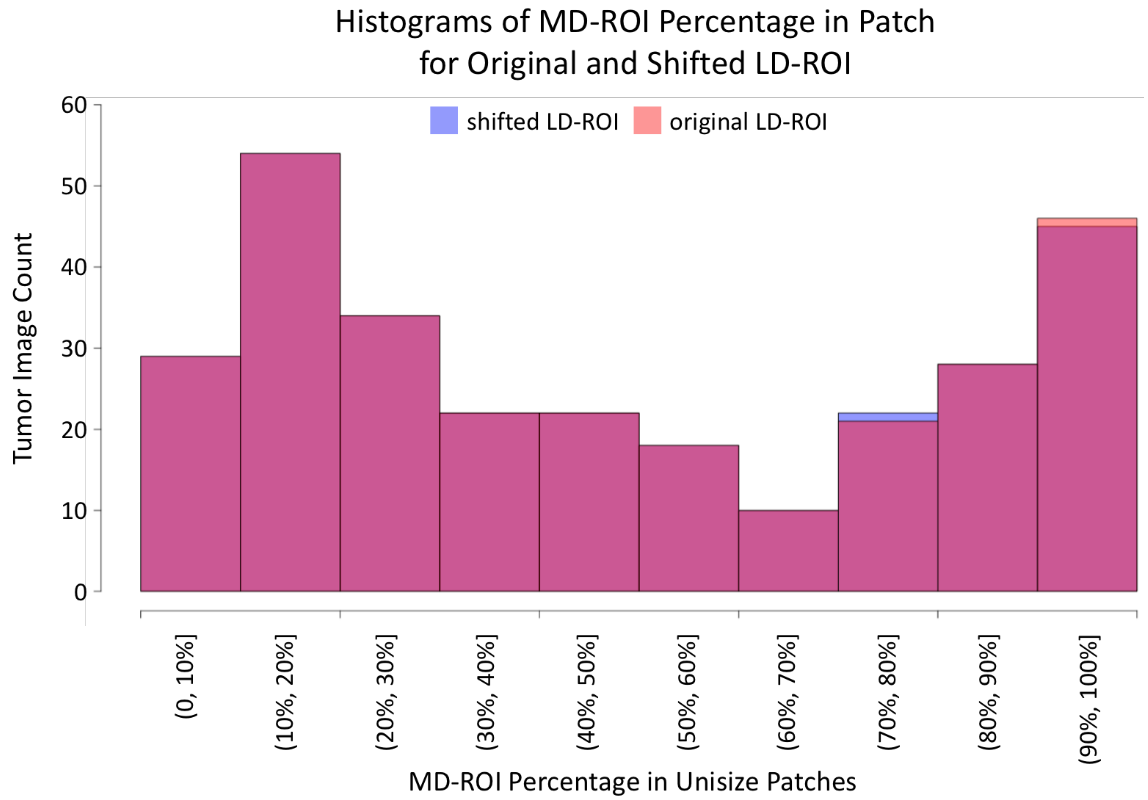


Figure 22: Comparison between Tumor Percentages for Original LD-ROIs and Shifted LD-ROIs

4.5 Conclusion

In this paper, we proposed a dual-representation transfer learning model with CNN for computer-aided diagnosis on breast tumors with CEDM images. We showed that full-tuning strategy is a necessary way to improve the performance on deep transfer learning architectures. The dual-representation approach we designed successfully increased the capabilities of the model for multi-modality medical images. Moreover, we concluded that a mid-size lazy-drawn ROI size (256×256) is the optimal selection which got the best performance on CEDM data. We also observed limited robustness of our model that with shifted LD-ROI, the sensitivity and specificity are not balanced but the accuracy remains.

Since the deep learning techniques are still new to the medical images research, especially to mammography topics, we believe there are several future directions we should follow up: 1) in our current work, we considered only InceptionV3 model, other well-designed models can be applied; 2) fixed LD-ROI helped reduce the time of labeling, a dynamic optimal value of ROI size is much needed for different applications; 3) with very limited data, data completion and augmentation techniques should be considered to improve the performance.

REFERENCES

Abujudeh, H. H., R. Kaewlai, B. A. Asfaw and J. H. Thrall, "Quality initiatives: key performance indicators for measuring and improving radiology department performance", Radiographics **30**, 3, 571–580 (2010).

Ahmed, T., R. Boutaba and A. Mehaoua, "A measurement-based approach for dynamic qos adaptation in diffserv networks", Computer Communications **28**, 18, 2020–2033 (2005).

Almes, G., S. Kalidindi and M. Zekauskas, "A one-way delay metric for IPPM", Tech. rep. (1999a).

Almes, G., S. Kalidindi and M. Zekauskas, "A one-way packet loss metric for IPPM", Tech. rep. (1999b).

American Cancer Society, "American cancer society recommendations for the early detection of breast cancer", URL https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html (2017).

Amoah, A., S. Amirfar, K. Sebek, S. L Silfen, J. Singer and J. Wang, "Developing composite quality measures for ehr-enabled primary care practices in new york city", **30**, 231–9 (2015).

Anderson, E., Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney *et al.*, *LAPACK Users' guide* (SIAM, 1999).

Arevalo, J., F. A. González, R. Ramos-Pollán, J. L. Oliveira and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks", Computer methods and programs in biomedicine **127**, 248–257 (2016).

Armato, S. G. and W. F. Sensakovic, "Automated lung segmentation for thoracic ct: Impact on computer-aided diagnosis", Academic Radiology **11**, 9, 1011–1021 (2004).

Association, N. E. M. *et al.*, "Digital imaging and communication in medicine (dicom)", NEMA PS 3 Supplement 23 Structured Reporting (1997).

Baker, S. D. and D. H. Hoglund, "Medical-grade, mission-critical wireless networks [designing an enterprise mobility solution in the healthcare environment]", IEEE Engineering in Medicine and Biology Magazine **27**, 2 (2008).

Berg, W. A., Z. Zhang, D. Lehrer, R. A. Jong, E. D. Pisano, R. G. Barr, M. Böhm-Vélez, M. C. Mahoney, W. P. Evans, L. H. Larsen *et al.*, "Detection of breast cancer with addition of annual screening ultrasound or a single screening mri to mammography in women with elevated breast cancer risk", Jama **307**, 13, 1394–1404 (2012).

Bersimis, S., S. Psarakis and J. Panaretos, "Multivariate statistical process control charts: an overview", Quality and Reliability engineering international **23**, 5, 517–543 (2007).

Blackford, S., "LAPACK benchmark", URL http://www.netlib.org/lapack/lug/node71.html (1999).

Cáceres, R., N. G. Duffield, J. Horowitz, D. Towlsey and T. Bu, "Multicast-based inference of network-internal characteristics: Accuracy of packet loss estimation", in "INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE", vol. 1, pp. 371–379 (IEEE, 1999).

Capizzi, G. and G. Masarotto, "A least angle regression control chart for multidimensional data", Technometrics **53**, 3, 285–296 (2011).

Cascarano, N., L. Ciminiera and F. Risso, "Optimizing deep packet inspection for high-speed traffic analysis", Journal of Network and Systems Management **19**, 1, 7–31 (2011).

Centers for Medicare & Medicaid Services (CMS), H. *et al.*, "Medicare and medicaid programs; electronic health record incentive program. final rule.", Federal register **75**, 144, 44313 (2010).

Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", arXiv preprint arXiv:1606.00915 (2016).

Cheng, J.-Z., D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans", Scientific reports **6**, 24454 (2016).

Ciavattone, L., A. Morton and G. Ramachandran, "Standardized active measurements on a tier 1 ip backbone", IEEE Communications Magazine **41**, 6, 90–97 (2003).

Cisco System, Inc., "Voice over IP - per call bandwidth consumption (document ID: 7934)", URL http://www.cisco.com/c/en/us/support/docs/voice/voice-quality/7934-bwidth-consume.html (2016).

Conway, A. E., "A passive method for monitoring voice-over-ip call quality with itu-t objective speech quality measurement methods", in "Communications, 2002. ICC 2002. IEEE International Conference on", vol. 4, pp. 2583–2586 (IEEE, 2002).

DeSimone, A. K., A. Post, R. Duszak Jr and P.-A. T. Duong, "Radiology trainee vs faculty radiologist fluoroscopy time for imaging-guided procedures: A retrospective study of 17,966 reports over a 5.5-year period", Current problems in diagnostic radiology (2017).

Dhungel, N., G. Carneiro and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests", in "Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on", pp. 1–8 (IEEE, 2015a).

Dhungel, N., G. Carneiro and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms", in "International Conference on Medical Image Computing and Computer-Assisted Intervention", pp. 605–612 (Springer, 2015b).

Doi, K., "Computer-aided diagnosis in medical imaging: historical review, current status and future potential", Computerized medical imaging and graphics **31**, 4, 198–211 (2007).

Efron, B., T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression", The Annals of statistics **32**, 2, 407–499 (2004).

Erl, T., *SOA: principles of service design*, vol. 1 (Prentice Hall Upper Saddle River, 2008).

Fraleigh, C., S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely and S. C. Diot, "Packet-level traffic measurements from the sprint ip backbone", IEEE network **17**, 6, 6–16 (2003).

Giger, M. L., N. Karssemeijer and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer", Annual review of biomedical engineering **15**, 327–357 (2013).

Haralick, R. M., K. Shanmugam *et al.*, "Textural features for image classification", IEEE Transactions on systems, man, and cybernetics , 6, 610–621 (1973).

Hartman, M., A. B. Martin, D. Lassman, A. Catlin, N. H. E. A. Team *et al.*, "National health spending in 2013: growth slows, remains in step with the overall economy", Health Affairs **34**, 1, 150–160 (2015).

He, K., X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition”, in “Proceedings of the IEEE conference on computer vision and pattern recognition”, pp. 770–778 (2016).

Hu, L. S., S. Ning, J. M. Eschbacher, N. Gaw, A. C. Dueck, K. A. Smith, P. Nakaji, J. Plasencia, S. Ranjbar, S. J. Price *et al.*, “Multi-parametric mri and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma”, PloS one **10**, 11, e0141506 (2015).

Hu, M., W. Pavlicek, P. T. Liu, M. Zhang, S. G. Langer, S. Wang, V. Place, R. Miranda and T. T. Wu, “Informatics in radiology: Efficiency metrics for imaging device productivity”, Radiographics **31**, 2, 603–616 (2011).

Inmon, W. H., *Building the data warehouse* (John wiley & sons, 2005).

International Telecommunication Union, “General aspects of quality of service and network performance in digital networks, including ISDNs (recommendation id: I.350)”, Tech. rep. (1993).

Jiang, W., K. Wang and F. Tsung, “A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis”, Journal of Quality Technology **44**, 3, 209 (2012).

Jong, R. A., M. J. Yaffe, M. Skarpathiotakis, R. S. Shumak, N. M. Danjoux, A. Gunesekara and D. B. Plewes, “Contrast-enhanced digital mammography: initial clinical experience”, Radiology **228**, 3, 842–850 (2003).

Joo, S., Y. S. Yang, W. K. Moon and H. C. Kim, “Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features”, IEEE transactions on medical imaging **23**, 10, 1292–1300 (2004).

Kaklamanos, I. G., K. Birbas, K. N. Syrigos, D. Vlachodimitropoulos, N. Goutas and G. Bonatsos, “Breast angiosarcoma that is not related to radiation exposure: a comprehensive review of the literature”, Surgery today **41**, 2, 163–168 (2011).

Kaye, D., *Loosely coupled: the missing pieces of Web services* (RDS Strategies LLC, 2003).

Kerlikowske, K., P. A. Carney, B. Geller, M. T. Mandelson, S. H. Taplin, K. Malvin, V. Ernster, N. Urban, G. Cutter, R. Rosenberg *et al.*, “Performance of screening mammography among women with and without a first-degree relative with breast cancer”, Annals of Internal Medicine **133**, 11, 855–863 (2000).

Kohn, L. T., J. Corrigan and M. S. Donaldson, “Institute of medicine (us). committee on quality of health care in america.”, Crossing the quality chasm: a new health system for the 21st century (2001).

Lévy, D. and A. Jain, “Breast mass classification from mammograms using deep convolutional neural networks”, arXiv preprint arXiv:1612.00542 (2016).

Liao, S., Y. Gao, A. Oto and D. Shen, “Representation learning: a unified deep learning framework for automatic prostate mr segmentation”, in “International Conference on Medical Image Computing and Computer-Assisted Intervention”, pp. 254–261 (Springer, 2013).

Liu, S., S. Liu, W. Cai, S. Pujol, R. Kikinis and D. Feng, “Early diagnosis of alzheimer’s disease with deep learning”, in “Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on”, pp. 1015–1018 (IEEE, 2014).

McVeigh, F. L., A. K. Tarbett, A. M. Betts and T. R. Boal, “Efficiency of automation and electronic health records in optometric practice”, Optometry-Journal of the American Optometric Association **79**, 1, 43–49 (2008).

Mei, Y., “Efficient scalable schemes for monitoring a large number of data streams”, Biometrika **97**, 2, 419–433 (2010).

Middleton, B., M. Bloomrosen, M. A. Dente, B. Hashmat, R. Koppel, J. M. Overhage, T. H. Payne, S. T. Rosenbloom, C. Weaver and J. Zhang, “Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from amia”, Journal of the American Medical Informatics Association **20**, e1, e2–e8 (2013).

Miller, R. H., C. West, T. M. Brown, I. Sim and C. Ganchoff, “The value of electronic health records in solo or small group practices”, Health Affairs **24**, 5, 1127–1137 (2005).

Moseley, M., “Eliminating data warehouse pressures with master data services and soa”, Business Intelligence Journal **14**, 2, 36 (2009).

Nagy, P. G., M. J. Warnock, M. Daly, C. Toland, C. D. Meenan and R. S. Mezrich, “Informatics in radiology: automated web-based graphical dashboard for radiology operational business intelligence”, Radiographics **29**, 7, 1897–1906 (2009).

Nitrosi, A., M. Bertolini, P. Notari, A. Botti, V. Ginocchi, G. Tondelli, M. Iori and P. Pattacini, “Efficiency and effectiveness of an innovative ris function for patient information reconciliation directly integrated with pacs”, Journal of digital imaging **26**, 3, 412–418 (2013).

Ojala, T., M. Pietikainen and D. Harwood, “Performance evaluation of texture measures with classification based on kullback discrimination of distributions”, in “Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on”, vol. 1, pp. 582–585 (IEEE, 1994).

Ondategui-Parra, S., J. G. Bhagwat, K. H. Zou, E. Nathanson, I. E. Gill and P. R. Ros, “Use of productivity and financial indicators for monitoring performance in academic radiology departments: Us nationwide survey”, Radiology **236**, 1, 214–219 (2005).

Oquab, M., L. Bottou, I. Laptev and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks”, in “Proceedings of the IEEE conference on computer vision and pattern recognition”, pp. 1717–1724 (2014).

Patel, B. K., R. J. Gray and B. A. Pockaj, “Potential cost savings of contrast-enhanced digital mammography”, American Journal of Roentgenology **208**, 6, W231–W237 (2017).

Paxson, V., J. Mahdavi, A. Adams and M. Mathis, “An architecture for large scale internet measurement”, IEEE Communications Magazine **36**, 8, 48–54 (1998).

Pearson, K., “Liii. on lines and planes of closest fit to systems of points in space”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**, 11, 559–572 (1901).

Peer, P. G., A. L. Verbeek, H. Straatman, J. H. Hendriks and R. Holland, “Age-specific sensitivities of mammographic screening for breast cancer”, Breast cancer research and treatment **38**, 2, 153–160 (1996).

Poissant, L., J. Pereira, R. Tamblyn and Y. Kawasumi, “The impact of electronic health records on time efficiency of physicians and nurses: a systematic review”, Journal of the American Medical Informatics Association **12**, 5, 505–516 (2005).

Roehr, B., “Health care in us ranks lowest among developed countries, commonwealth fund study shows”, Bmj **337**, 193 (2008).

Roesch, M. *et al.*, “Snort: Lightweight intrusion detection for networks.”, in “Lisa”, vol. 99, pp. 229–238 (1999).

Rosenblatt, F., *The perceptron, a perceiving and recognizing automaton Project Para* (Cornell Aeronautical Laboratory, 1957).

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, “ImageNet large scale visual recognition challenge”, International Journal of Computer Vision **115**, 3, 211–252 (2015).

Sarwar, A., G. Boland, A. Monks and J. B. Kruskal, “Metrics for radiologists in the era of value-based health care delivery”, Radiographics **35**, 3, 866–876 (2015).

Schwedt, T. J., C. D. Chong, T. Wu, N. Gaw, Y. Fu and J. Li, “Accurate classification of chronic migraine via brain magnetic resonance imaging”, Headache: The Journal of Head and Face Pain **55**, 6, 762–777 (2015).

Seltzer, S. E., P. Kelly, G. M. Deibel and P. Ros, “Radiology quality and performance metrics on the web: a management information and communications tool”, Academic radiology **7**, 11, 981–985 (2000).

Shin, H.-C., M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, “Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data”, IEEE transactions on pattern analysis and machine intelligence **35**, 8, 1930–1943 (2013).

Simonyan, K. and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556 (2014).

Smith, R., C. Estan, S. Jha and S. Kong, “Deflating the big bang: fast and scalable deep packet inspection with extended finite automata”, in “ACM SIGCOMM Computer Communication Review”, vol. 38, pp. 207–218 (ACM, 2008).

Stallings, W. and C. T, *Business Data Communications: Infrastructure, Networking and Security* (Pearson, 2013).

Suk, H.-I., S.-W. Lee, D. Shen, A. D. N. Initiative *et al.*, “Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis”, NeuroImage **101**, 569–582 (2014).

Suk, H.-I. and D. Shen, “Deep learning-based feature representation for ad/mci classification”, in “International Conference on Medical Image Computing and Computer-Assisted Intervention”, pp. 583–590 (Springer, 2013).

Sun, T., R. Zhang, J. Wang, X. Li and X. Guo, “Computer-aided diagnosis for early-stage lung cancer based on longitudinal and balanced data”, PloS one **8**, 5, e63559 (2013).

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, “Going deeper with convolutions”, in “Proceedings of the IEEE conference on computer vision and pattern recognition”, pp. 1–9 (2015).

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, “Rethinking the inception architecture for computer vision”, in “Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition”, pp. 2818–2826 (2016).

Tartakovsky, A. G. and V. V. Veeravalli, “Asymptotically optimal quickest change detection in distributed sensor systems”, Sequential Analysis **27**, 4, 441–475 (2008).

Tibshirani, R., “Regression shrinkage and selection via the lasso”, Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996).

van Ginneken, B., C. M. Schaefer-Prokop and M. Prokop, “Computer-aided diagnosis: how to move from the laboratory to the clinic”, Radiology **261**, 3, 719–732 (2011).

Walker, J. M., P. Carayon, N. Leveson, R. A. Paulus, J. Tooker, H. Chin, A. Bothe Jr and W. F. Stewart, “EHR safety: the way forward to safe and effective systems”, Journal of the American Medical Informatics Association **15**, 3, 272–277 (2008).

Wang, D. C., C. R. Parry, M. Feldman, G. Tomlinson, J. Sarrazin and P. Glanc, “Acute abdomen in the emergency department: is ct a time-limiting factor?”, American Journal of Roentgenology **205**, 6, 1222–1229 (2015).

Wang, K. and W. Jiang, “High-dimensional process monitoring and fault isolation via variable selection”, Journal of Quality Technology **41**, 3, 247 (2009).

Wang, K., C. Zwart, C. Wellnitz, T. Wu and J. Li, “Integration of multiple health information systems for quality improvement of radiologic care”, IISE Transactions on Healthcare Systems Engineering **7**, 3, 169–180 (2017).

Wang, S., W. Pavlicek, C. C. Roberts, S. G. Langer, M. Zhang, M. Hu, R. L. Morin, B. A. Schueler, C. V. Wellnitz and T. Wu, “An automated dicom database capable of arbitrary data mining (including radiation dose indicators) for quality monitoring”, Journal of digital imaging **24**, 2, 223–233 (2011).

Wang, S. J., B. Middleton, L. A. Prosser, C. G. Bardon, C. D. Spurr, P. J. Carchidi, A. F. Kittler, R. C. Goldszer, D. G. Fairchild, A. J. Sussman *et al.*, “A cost-benefit analysis of electronic medical records in primary care”, The American journal of medicine **114**, 5, 397–403 (2003).

Warner, E., D. B. Plewes, K. A. Hill, P. A. Causer, J. T. Zubovits, R. A. Jong, M. R. Cutrara, G. DeBoer, M. J. Yaffe, S. J. Messner *et al.*, “Surveillance of brca1 and brca2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination”, Jama **292**, 11, 1317–1325 (2004).

Williamson, B., *Developing IP multicast networks*, vol. 1 (Cisco Press, 2000).

Wu, L., G. Barash and C. Bartolini, “A service-oriented architecture for business intelligence”, in “Service-Oriented Computing and Applications, 2007. SOCA’07. IEEE International Conference on”, pp. 279–285 (IEEE, 2007).

Xu, Y., T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric and C. Chang, “Deep learning of feature representation with multiple instance learning for medical image analysis”, in “Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on”, pp. 1626–1630 (IEEE, 2014).

Yu, F., Z. Chen, Y. Diao, T. Lakshman and R. H. Katz, “Fast and memory-efficient regular expression matching for deep packet inspection”, in “Architecture for Networking and Communications systems, 2006. ANCS 2006. ACM/IEEE Symposium on”, pp. 93–102 (IEEE, 2006).

Zeiler, M. D. and R. Fergus, “Visualizing and understanding convolutional networks”, in “European conference on computer vision”, pp. 818–833 (Springer, 2014).

Zeng, T. and S. Ji, “Deep convolutional neural networks for multi-instance multi-task learning”, in “Data Mining (ICDM), 2015 IEEE International Conference on”, pp. 579–588 (IEEE, 2015).

Zhang, J., “Powerful goodness-of-fit tests based on the likelihood ratio”, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**, 2, 281–294 (2002).

Zhang, W., R. Li, T. Zeng, Q. Sun, S. Kumar, J. Ye and S. Ji, “Deep model based transfer and multi-task learning for biological image analysis”, IEEE Transactions on Big Data (2016).

Zou, C. and P. Qiu, “Multivariate statistical process control using LASSO”, Journal of the American Statistical Association **104**, 488, 1586–1596 (2009).

Zou, C., Z. Wang, X. Zi and W. Jiang, “An efficient online monitoring method for high-dimensional data streams”, Technometrics **57**, 3, 374–387 (2015).

Zou, H., “The adaptive lasso and its oracle properties”, Journal of the American statistical association **101**, 476, 1418–1429 (2006).

Zou, H., T. Hastie, R. Tibshirani *et al.*, “On the “degrees of freedom” of the LASSO”, The Annals of Statistics **35**, 5, 2173–2192 (2007).

APPENDIX A

BRIEF INTRODUCTION TO SINGULAR VALUE DECOMPOSITION

SVD is a matrix factorization method. Let $\mathbf{Z}$ be a $i \times j$ matrix. The SVD of $\mathbf{Z}$ is:

$$\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^T = \sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^T \tag{A.1}$$

where r is the rank of $\mathbf{Z}$, $\mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_r)$ is a matrix consisting of orthonormal left singular vectors, $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_r)$ is a matrix consisting of orthonormal right singular vectors, $\mathbf{S}$ is a diagonal matrix with positive singular values $s_1 \geq \cdots \geq s_r$ on its diagonal. $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$ can be obtained by eigendecomposition. A.1 indicates that SVD decomposes $\mathbf{Z}$ into a summation of r rank-one matrices, $\mathbf{Z}_k = s_k \mathbf{u}_k \mathbf{v}_k^T$, $k = 1, \cdots, r$. It has been shown that $\mathbf{Z}_1$ is the closest rank-one approximation to $\mathbf{Z}$ in terms of minimizing the square Frobenius norm. This means that s_1, $\mathbf{u}_1$, $\mathbf{v}_1$ can be obtained by solving the following optimization problem:

$$\begin{gathered}(\hat{s}_1, \hat{\mathbf{u}}_1, \hat{\mathbf{v}}_1) = \underset{s_1, \mathbf{u}_1, \mathbf{v}_1}{\arg\min} \left\| \mathbf{Z} - s_1 \mathbf{u}_1 \mathbf{v}_1^T \right\|_F^2 \\ s.t. \left\| \mathbf{u}_1 \right\|^2 = 1, \left\| \mathbf{v}_1 \right\|^2 = 1, s_1 \geq 0\end{gathered}$$

APPENDIX B

PROOF OF PROPOSITION 3.1

It is easy to show that $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$ is a rank-one matrix. Therefore,

$$E[\mathbf{X}] - np_0\mathbf{1}_{q\times m} = \sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^T \tag{B.1}$$

In other words, $s_k = 0$ for $k > 1$. Furthermore, let $\{E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}\}_{ij,t}$ be the element of matrix $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$ at the row corresponding to the sender-receiver pair (i,j) and the t-th column. We know from the definition of $E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}$ that

$$\{E[\mathbf{X}] - np_0\mathbf{1}_{q\times m}\}_{ij,t} = \begin{cases} 0 & \text{for } (i,j) \notin \mathcal{F} \\ -n\delta(t) & \text{for } (i,j) \in \mathcal{F} \end{cases} \tag{B.2}$$

Using Equation B.1, we can further write Equation B.2 into

$$s_1 u_{1,ij} v_{1,t} = \begin{cases} 0 & \text{for } (i,j) \notin \mathcal{F} \\ -n\delta(t) & \text{for } (i,j) \in \mathcal{F} \end{cases} \tag{B.3}$$

where $u_{1,ij}$ is the element of $\mathbf{u}_1$ corresponding to the sender-receiver pair (i,j) and $v_{1,t}$ is the the t-th element of $\mathbf{v}_1$.

Let (i',j') be a sender-receiver pair that is affected by the fault and $(\hat{i},\hat{j})$ be one that is not, i.e., $(i',j') \in \mathcal{F}$ and $(\hat{i},\hat{j}) \notin \mathcal{F}$. Then, according to Equation B.3,

$$s_1 u_{1,\hat{i}\hat{j}} v_{1,t} = 0 \text{ and } s_1 u_{1,i'j'} v_{1,t} \neq 0 \tag{B.4}$$

since $s_1 \neq 0$ by the definition of SVD. Then, Equation B.4 becomes

$$u_{1,\hat{i}\hat{j}} v_{1,t} = 0 \text{ and } u_{1,i'j'} v_{1,t} \neq 0 \tag{B.5}$$

The sufficient and necessary condition for the above simultaneous equations to hold is $v_{1,t} \neq 0$, $u_{1,\hat{i}\hat{j}} = 0$, and $u_{1,i'j'} \neq 0$. Next, we derive the formula for $u_{1,i'j'}$, $v_{1,t}$, and s_1.

Let (i'',j'') be another sender-receiver pair that is affected by the fault, i.e., $(i'',j'') \in \mathcal{F}$. According Equation B.3, $s_1 u_{1,i'j'} v_{1,t} = s_1 u_{1,i''j''} v_{1,t} = -n\delta(t)$, i.e., $u_{1,i'j'} v_{1,t} = u_{1,i''j''} v_{1,t}$. Furthermore, because $\mathbf{u}_1$ is orthonormal, we have $\sum_{(i,j)\in mathF} u_{1,ij}^2 = |\mathcal{F}| \times u_{1,ij}^2 = 1$. Solving this equation gives $u_{1,ij} = \frac{1}{\sqrt{|\mathcal{F}|}}$ for $\forall (i,j) \in \mathcal{F}$.

To derive the formula for $v_{1,t}$ and s_1, focus on a sender-receiver pair $(i,j) \in \mathcal{F}$. Then, $s_1 u_{1,ij} v_{1,t} = s_1 \frac{1}{\sqrt{|\mathcal{F}|}} v_{1,t} = -n\delta(t)$, i.e.,

$$v_{1,t} = \frac{-n\delta(t)\sqrt{|\mathcal{F}|}}{s_1} \tag{B.6}$$

Using the property that $\mathbf{v}_1$ is orthonormal, we have $\sum_t v_{1,t}^2 = 1$. Combining this with Equation B.6, we get $s_1 = \sqrt{\sum_{t=1}^{m} [n\delta(t)]^2} \times \sqrt{|\mathcal{F}|}$. Inserting this into Equation B.6, we get $v_{1,t} = \frac{-\delta(t)}{\sum_{t=1}^{m} \delta^2(t)}$.