DOCTOR OF PHILOSOPHY

Cloud enabled data analytics and visualization framework for health-shock prediction

Mahmud, Shahid

Award date:
2016

Awarding institution:
Coventry University

Cloud Enabled Data Analytics and Visualization Framework for Health-Shock Prediction

By

Shahid Mahmud

March 2016

Coventry University

Cloud Enabled Data Analytics and Visualization Framework for Health-Shock Prediction

By

Shahid Mahmud

March 2016

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

Abstract

Health-shock can be defined as a health event that causes severe hardship to the household because of the financial burden for healthcare payments and the income loss due to inability to work. It is one of the most prevalent shocks faced by the people of underdeveloped and developing countries. In Pakistan especially, policy makers and healthcare sector face an uphill battle in dealing with health-shock due to the lack of a publicly available dataset and an effective data analytics approach. In order to address this problem, this thesis presents a data analytics and visualization framework for health-shock prediction based on a large-scale health informatics dataset. The framework is developed using cloud computing services based on Amazon web services integrated with Geographical Information Systems (GIS) to facilitate the capture, storage, indexing and visualization of big data for different stakeholders using smart devices. The data was collected through offline questionnaires and an online mobile based system through Begum Memhooda Welfare Trust (BMWT). All data was coded in the online system for the purpose of analysis and visualization. In order to develop a predictive model for health-shock, a user study was conducted to collect a multidimensional dataset from 1000 households in rural and remotely accessible regions of Pakistan, focusing on their health, access to health care facilities and social welfare, as well as economic and environmental factors.

The collected data was used to generate a predictive model using a fuzzy rule summarization technique, which can provide stakeholders with interpretable linguistic rules to explain the causal factors affecting health-shock. The evaluation of the proposed system in terms of the interpretability and accuracy of the generated data models for classifying health-shock shows promising results. The prediction accuracy of the fuzzy model based on a k-fold cross-validation of the data samples shows above 89% performance in predicting health-shock based on the given factors. Such a framework will not only help the government and policy makers to manage and mitigate health-shock effectively and timely, but will also provide a low-cost, flexible, scalable, and secure architecture for data analytics and visualization.

Future work includes extending this study to form Pakistan's first publicly available health informatics tool to help government and healthcare professionals to form policies and healthcare reforms. This study has implications at a national and international level to facilitate large-scale health data analytics through cloud computing in order to minimize the resource commitments needed to predict and manage health-shock.

Acknowledgement

I would like to express my sincere gratitude to my Director of Studies, Dr. Rahat Iqbal and my Supervisor, Dr. Faiyaz, for their invaluable guidance, excellent suggestions, constant encouragement, and kind understanding during the dissertation research period. I would also like to express my sincere gratitude to Prof. Babak Akhgar and Dr. Nazaraf Shah for their invaluable time and efforts to review this dissertation and provide me with their insightful comments.

I would also like to take this opportunity to thank local charitable trust, namely; Begum Memhooda Welfare Trust (BMWT) and the Quaid-e-Azam University (QAU), for their help and support in data collection.

I am grateful to Muhammad Rehman Zafar for his support in developing a purpose built mobile application for survey activity. I thank you all for your unconditional support and valued assistance.

For all others who, more or less, helped me in this work, please be assured that I am forever grateful to you all.

Last but not the least, I would like to express my heartiest gratitude and love to my family for their great love, continuous support, and encouragement during my graduate studies.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Description
DEWS	Disease Early Warning System
MMR	Maternal Mortality Rate
IMR	Infant Mortality Rate
ROC	Receiver Operating Curve
MPAT	Multidimensional Poverty Assessment Tool
MDGs	Millennium Development Goals
PSP	Poverty Scorecard for Pakistan
PPAF	Pakistan Poverty Alleviation Fund
GP	General Practitioners
NGOs	Non-Governmental Organizations
WHO	World Health Organization
HHF	Heartfile Health Financing
TFR	Total Fertility Rate
ASR	Adult Survival Rates
HIIF	Health Information Integration Framework
ICT	Information and Communication Technology
EHR	Electronic Health Records
SaaS	Software as a Service
PaaS	Platform as a Service)
IaaS	Infrastructure as a Service
IoT	Internet of Things
SES	Socio Economic Status
BMWT	Begum Mehmooda Welfare Trust
EBS	Elastic Block Storage

List of Publications

Journal Publications

1. **Mahmud, S.**, Iqbal, R., and Doctor, F. (2015). Cloud Enabled Data Analytics and Visualization Framework for Health Shock Prediction. *Future Generation Computer Systems*, http://dx.doi.org/10.1016/j.future.2015.10.014, December.

Conference Publications

2. **Mahmud, S.**, Iqbal, R., and Doctor, F. (2014). An integrated framework for the prediction of health shocks. *In Proceedings of The 2nd International Conference on Applied Information and Communications Technology*, ICAICT.
3. **Mahmud, S.**, Iqbal, R., and Doctor, F. (2015). Context -Aware HealthCare Dataset - A case study from Pakistan. *In The Fifth International Conference on Advances in Information Mining and Management*, IMMM, pages 77–82.

Chapter 1

Introduction

1.1 Introduction

This chapter presents background information and discusses the research aims and objectives. Furthermore, it presents some concepts related to health-shock in Pakistan. The motivation for this work is also discussed in this chapter.

The rest of the chapter is organized as follows: Section 1.2 presents a background of health, healthcare system and health-shock in addition to some problems related to this area of research; Section 1.3 describes the research motivation in brief; Section 1.4 presents the problem statement; Section 1.5 presents the aims and objectives; Section 1.6 outlines the research scope; Section 1.7 presents the research methodology; Section 1.8 outlines the main contributions of this thesis; and Section 1.9 presents the structure of the thesis.

1.2 Background

Health is a potential state in which someone performs well, both mentally and physically living within one's inhabiting environment (WHO, 2014; William, 1994). Likewise, the provision of health services by utilizing available health resources to fulfill the needs of a target population is termed a healthcare delivery system (William, 1994). In the progress of a country, quality of healthcare services plays a vital role and depends greatly upon the socio-economic conditions of the region and the people living there. In this regard, Pakistan is placed 6^{th} among the densely populated states of the world having a huge concentration living in rural areas (Mahmud et al., 2014; Nishtar et al., 2013). Therefore, in order to meet the challenges in providing health services evenly in rural and urban areas, Pakistan is striving to achieve the Millennium Development Goals (MDGs) (Naveed and Islam, 2012; WHO, 2015). Here, it is worth mentioning that by analyzing the performance of the past decade, it reveals that the health sector has been deprived of the required attention while education and security sectors have remained on the front-line.

Furthermore, inadequate policies and reforms at the government level has caused health-care delivery to deteriorate, specifically in the rural areas (Nishtar, 2011; Nishtar et al., 2013). Moreover, government health facilities are insufficient due to the lack of much needed resources (Nishtar, 2011; Nishtar et al., 2013). Because of the current situation, people seek help from private service providers at a much higher cost, especially in under-developed areas of Pakistan (Mahmud et al., 2015; Nishtar, 2011). Furthermore, abrupt changes in socio-economic situations have increased uncertainty levels causing constant anxiety and stress among people (Mahmud et al., 2015). As a result, poverty triggers the neglect of health and the corresponding vicious cycle of health-shock (Alam, 2009; Balasubramanian, 2008). Here, it is worth mentioning that in Pakistan, health-shock is more concentrated amongst the poorer population given the higher prevalence of diseases in remote areas (Pal and Palacios, 2011; Schreiner, 2010).

Generally, health-shock is defined as the unpredictable illness of the principal breadwinner in the family and its social and economic effect on the individual, family, society, and various governance levels (Alam and Mahal, 2014; Batts, 2001; Leive and Xu, 2008; Mahmud et al., 2014; Weitz, 1980). Health-shock is one of the most common and severe shocks faced by households in developing countries (Dreze and Srinivasan, 1997; Gupta and Sankar, 2001; Leive and Xu, 2008). These households struggle with incomes that are low and volatile. The extent and duration of health shock can make it very difficult for a household to cope and may cause food insecurity, irreversible malnutrition, termination of schooling, and/or breaking of matrimonial links.

Health-shock prevents possible participation of individuals in the labour force, reducing productivity and hours of work, hence reducing earnings (Gupta and Sankar, 2001; Leive and Xu, 2008). Unhealthy individuals, especially the elderly, become a liability and burden for the household, and take away large part of household income for their medical expenses. Moreover, it prompts households to engage in low return activities which hinder the ability of the poor to grow their incomes and escape the poverty cycle (Elbers et al., 2007; Rosenzweig and Binswanger, 1993). The aftermath of health-shock forces the household to deal with financial crises for a long time, possibly for their entire life (Baldwin et al., 1994; Burkhauser and Daly, 1996; Burkhauser and Mary, 1994).

In order to predict and mitigate health-shock, the main challenges include:

1. Capturing, integrating and categorizing different data variables pertaining to health, social/welfare, environment and economic indicators contributing to health-shock.

2. Analysis of the data through the use of appropriate supervised and unsupervised machine learning and data mining techniques to visualize and predict the country's health-shock landscape and identify factors and trends affecting the risk of health-shock. Effective analysis of such datasets can provide better management and control of epidemics and critical illness.

1.3 Motivation

As per "Human Development Report" released by the United Nations Development Programme (UNDP) in 2014 (UNDP, 2014), Pakistan's Human Development Indices (HDI) stood at 146^{th} position out of a total 187 countries with 60.19% of its population living below the poverty line, i.e., earning below USD 2/= per day. In this regard, a multi-dimensional poverty measurement survey has also been conducted by Benazir Income Support Programme (BISP) across the country using the Poverty Score Card (PSC) (GoP, 2012, 2015; Schreiner, 2010). PSC not only provides a poverty profile for each household but also provides data on 12 key indicators such as household size, type of housing and toilet facilities, education, child status, household assets, agricultural landholding, and livestock ownership. However, it does not include any indicator related to the health status of a household which was declared as one of the major contributors to economic shocks faced by families (GoP, 2013).

Generally, there is very slow progress and futile efforts being made on reforms for upgrading the legacy healthcare systems in Pakistan. Moreover, the currently deployed system in Pakistan, the Disease Early Warning System (DEWS) only focusses on reducing mortality and morbidity (WHO, 2008). Here, it is worth mentioning that DEWS does not focus on health-shock and its impact on a community (WHO, 2008).

Currently, in Pakistan, there is no dataset which can help to understand the root cause, or on which predictive modelling and data visualization can be applied in order to visualize, predict, and mitigate health-shock. Thus, this research is an attempt to collect a dataset for this purpose. The main question this thesis attempts at addressing is:

Is it possible to predict health-shock based on qualitative data analysis using artificial intelligence (AI) techniques?

1.4 Problem Statement

To predict is to save lives, especially in the domain of healthcare. The more accurately one can predict health related outcomes, the greater are the chances and opportunities to intervene, diagnose and implement pre-emptive measures, thus saving more lives as a result. Similarly, in order to develop a framework for prediction, it is extremely important to identify the factors that directly or indirectly give rise to health-shock. Furthermore, these factors depend greatly on the socio-economic, geographical, and cultural norms of that region so there is a dire need to conduct large-scale health surveys. However, due to the lack of publicly available datasets, there is no such prediction framework that can help in predicting and mitigating health-shock. Especially, in Pakistan, because, since the 18th amendment in the constitution of Pakistan (Nishtar et al., 2013), health has become a provincial responsibility making it more complex to define standards, or design a system at the federal level in order to monitor and mitigate health-shock.

1.5 Research Aims and Objectives

The aim of this research is to develop an intelligent model, which can help to predict health-shock and its effects at individual, family, and national level. To meet the above aim, the following objectives are defined:

- Review of the related literature including the state-of-the-art ICT approaches for healthcare management.
- Conduct an extensive user study on individuals and families residing in the rural and tribal areas of Pakistan in order to identify the main factors leading to health-shock in Pakistan.
- Carry out data analysis using statistical approaches to identify patterns and co-relations between data attributes for the prediction of health-shock.
- Develop a cost effective framework for data analysis and visualization.
- Carry out data modelling using fuzzy logic.
- Validate the developed model in order to determine the key indicators contributing to health-shock.

1.6 Research Scope

1. The scope of the research is to gather data from households living in the rural and tribal areas of Pakistan, which can help to investigate, formulate, simulate and model health-shock for the surveyed geographical area.

2. This research will then provide a set of recommendations for an eventual nationwide system and a pilot to enhance the work. (Note: This work is not intended to address or replace any existing country-wide health survey or country-wide scalable system to gauge and collect health related data from the masses.) Importantly, the research will serve as an alternative way to monitor and mitigate health-shock in a country where there is currently no such system.

1.7 Research Methodology

Research, regardless of the discipline or nature of it, entails either a qualitative or quantitative method, and/or a combination of both, for collecting and analyzing data (Creswell, 2013; Marwell, 2013). Generally, the qualitative approach focuses on describing a phenomenon or an event in a rather comprehensive fashion by means of comprehending the experiences, perspectives and thoughts of participants (Creswell, 2013; Marwell, 2013). Such an approach tends to be inductive, i.e., the theory/hypothesis, explanations and conceptualizations are all formulated on the basis of respondents' answers or perspectives (Saunders et al., 2011). Qualitative approaches facilitate the discovery and exploration of new phenomenon, providing an in-depth understanding of mechanisms, and facilitating the uncovering of data via surveys/questionnaires, just to name a few (Creswell, 2013).

On the other hand, the quantitative approach is primarily concerned with optimizing the objectivity, replicability, and generalizability of the data. The quantitative approach professes that the social reality can be objectively measured; whereby, accurate knowledge can be deduced (Bryman, 1996; Chalmers, 2004; Morgan and Smircich, 1980; Saunders et al., 1998). The quantitative research is a formal, systematic, and objective process, which helps in describing and testing the relationship and causal effects among variables. Advantages of the quantitative method include data collection from a large sample and generalisation of data (information and findings), which may be used for policy formation (Creswell, 2013).

Thereby, with regard to the above stated discourse of qualitative and quantitative approaches, it is quite evident that the quantitative method is viable due to the very nature and context of this particular study. Here, the quantitative method is suitable, because the research is concerned with improving the health of the population of people in the rural areas of Pakistan.

Furthermore, in order to address the research problems and challenges, a systematic methodology has been used as shown in Fig. 1.1. Details of the methodology are given below:

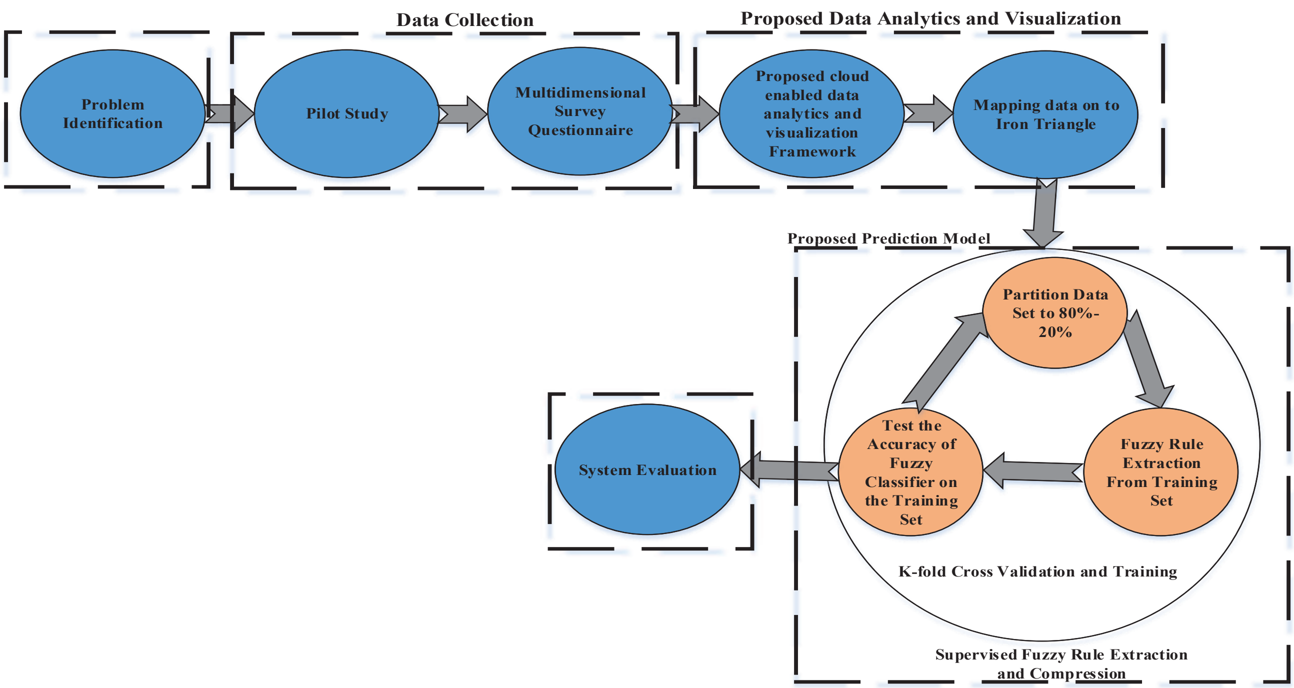


Figure 1.1: Research Methodology.

1. To gain an insight of socio-economic, health, cultural, and geographical norms of the people living in the rural areas of Pakistan, a survey has been conducted which also includes observing people while they answer questions in order to gain further insights.

2. As a part of data collection, a multi-dimensional survey questionnaire was distributed among 300 families living in the proximity of the Haripur district as a formal pilot study. Here, one of the main objectives was to further refine the survey questionnaire

based on the realities “on the ground” and the feedback of the respondents.

3. As a result of the pilot study and feedback of the respondents, a multi-dimensional survey questionnaire comprising of 47 features was fashioned and distributed among 1000 households. Some of these features include: the age, marital status, gender and involvement of the household head, involvement in the labour force, education of children, financial and water resources, access to health facilities, schools and clean water, effects of climatic changes (i.e. frequency of minor illness due to changes in weather, resistance of house against strong winds, severe rain, snow or hail), effects of shortage of basic facilities such as fuel, food, money for treatment of illness or fertilizer for crops and waste disposal trends.

4. Based on the collected dataset, a data visualization tool has been developed, which helps in visualization of the dataset for better understanding. Here, visualization can help various stakeholders in viewing the dataset according to their requirements and specialities.

5. Based on the visualization, mapping of the collected dataset on to the well-known iron triangle model of healthcare (William, 1994) has been performed.

6. To have an insight of financial resources such as costs, assets and the overall impact on people's health and accessibility to health facilities, the dataset also contains monetary values.

7. Finally, a cost effective, cloud enabled data analytics and visualization framework has been proposed for health-shock prediction. Here, the collected data has been used to generate a predictive model based on the fuzzy rule summarization technique. Furthermore, the proposed method has been evaluated in terms of interpretability and classification accuracy. Here, it is worth mentioning that the proposed prediction model shows a prediction accuracy of above 89% for predicting health-shock based on the given factors.

1.8 Research Contributions

The main contributions of this thesis are enumerated as follows:

1. Development of a context-aware healthcare dataset to understand and monitor health-shock in Pakistan. Based on the socio-economic, cultural, and geographic norms, a user study, based on a survey questionnaire, comprising of 47 features was conducted in the rural areas of Pakistan to collect the data from 1000 households using probability sampling. The questionnaire mainly focusses on poverty indicators and their relevance to the health impact indicators, unlike the Poverty Scorecard for Pakistan (PSP) and the Pakistan Poverty Alleviation Fund (PPAF) questionnaires; where the former only focuses upon the poverty dimension and the latter only focuses upon gauging the impact of PPAF activities for poverty alleviation (GoP, 2012, 2015; Schreiner, 2010).

2. Development of a cloud enabled mobile application with the aim of bringing efficiency in health data collection by eliminating data redundancy and errors associated with the data entry of the collected survey data. Furthermore, it enables the healthcare professionals to create and conduct large-scale health surveys with minimal human and financial resources.

3. Mapping of the collected dataset on to the well-known iron triangle model of healthcare (William, 1994), consisting of three dimensions; access, cost containment and high quality care. This depicts the socio-economic, health, cultural, and geographical norms of Pakistan.

4. Development of the data analysis and visualization tool aimed at understanding the hidden patterns in the data that through analysis can lead to policies and recommendations for better planning, cost efficiency, and improving the quality of life for the target population.

5. Development of a predictive model based on a fuzzy rule based summarization technique for predicting and validating health-shock. Furthermore, it is one of the first

comprehensive models that can predict health-shock based on qualitative data analysis. Here, the modelling of data was carried out using fuzzy rule based summarization approach.

1.9 Structure of the Thesis

The structure of this thesis is as follows:

- **CHAPTER 1: Introduction**

  This chapter presents an overview of the current research. It highlights the research importance and also gives a brief background about the problem domain. Moreover, it discusses the aim and objectives of the research, the research scope and the research contributions. The structure of the thesis is also discussed in this chapter.

- **CHAPTER 2: Research Background**

  This chapter provides a comprehensive insight into the research background of health-shock in Pakistan as shown through case study and the interrogation of various datasets. Furthermore, it discusses the involvement of Information and Communication Technology (ICT) in contemporary healthcare systems. One of the main objectives of this chapter is to review the existing approaches used in ICT based systems concerning factors relating to health, but more importantly, to investigate an artificial intelligent approach that can be used to analyze and model the data while taking into account uncertainties in the dataset.

- **CHAPTER 3: Literature Review**

  This chapter provides a detailed literature review on the existing works related to healthcare in underdeveloped and developing countries. Furthermore, it discusses the challenges which big data and cloud based systems bring to this field.

- **CHAPTER 4: Data Collection**

This chapter offers a detailed account of the dataset including the research instruments used for data collection, ethical considerations, and the pilot study. Furthermore, it discusses the details of the collected dataset and its importance, in order to understand the relationship between socio-economic and, demographic, and geographical factors and their impact on health.

- **CHAPTER 5: Cloud Enabled Data Analytics and Visualization Framework**

  This chapter presents the use of a cloud enabled framework for data analytics and visualization. The proposed framework is developed using cloud computing services based on Amazon web services (AWS) integrated with geographical information systems (GIS) to facilitate the capture, storage, indexing, and visualization of big data, through smart devices, for different stakeholders.

- **CHAPTER 6: Proposed Framework for Health−Shock Prediction**

  This chapter describes the framework for health-shock prediction in addition to the evaluation and validation of the respective model, i.e., performance and accuracy for predicting the health-shock. Furthermore, it presents the proposed Fuzzy summarization rule based approach which will help to predict health-shock based on the collected dataset.

- **CHAPTER 7: Conclusions and Future Work**

  This chapter summarizes the work presented in this thesis and outlines the directions for future research.

Chapter 2

Research background

2.1 Introduction

The chapter endeavours to describe and explain the background of health-shock and its impact in Pakistan. Furthermore, it discusses the impact of socio-economic, cultural, and geographical norms of any country on health-shock.

The schema of the chapter is as follows: Section 2.2 provides a discourse on healthcare in Pakistan by relating it to the "Iron Triangle Model of Healthcare" system; Section 2.3 discusses health-shock and its cause-and-effect in addition to its impact on people's health; Section 2.4 presents the case of Pakistan, discussing the experiences of households and the current health financing options available for poor people affected by health-shock; and Section 2.5 concludes the respective discourse with a summary.

2.2 Healthcare System

A healthcare system is the organization of people, institutions, and resources that deliver healthcare services to meet the needs of a target population (William, 1994). Generally, healthcare systems are evaluated based on three main factors: quality, cost, and accessibility to healthcare, also known as "The Iron Triangle of Healthcare" as shown in Fig. 2.1. In the case of an effective healthcare system, there should be a balance between all the three

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

Figure 2.1: The Iron Triangle Model of Healthcare (William, 1994).

components, i.e., the iron triangle should be an equilateral triangle, with each angle of 60° (William, 1994). However, in practice, any effective healthcare system can only optimize two of the three factors. For instance, to achieve higher access and quality, its associated cost will increase (Burns, 2014). Furthermore, these factors highly depend on the socio-economic, geographic, and cultural norms.

Pakistan has a dense hierarchical healthcare system when it comes to providing medical services to the nation. After the 18th amendment in the constitution of Pakistan, health has become a provincial responsibility. Here, different hierarchical structures at the provincial level in addition to the lack of a national health registry programme, have further complicated the healthcare system of Pakistan. Moreover, weak inter-provincial and inter-department linkages, growing population, medical malpractices, pilferage and the lack of a nation-wide integrated healthcare system, are acting as one of the major hindrances in policy making, evidence-based decision making, and in understanding the reasons behind health-shock.

2.3 Health-Shock

A detailed survey of health-shock and its comparison with different types of shocks including economical, natural, political, cultural, and religious, just to name a few has been done in (Doorslaer, 2006, 2007; Wagstaff and Doorslaer, 2003; Wagstaff and Pradhan, 2005). These surveys not only provide an insight to health-shock and its after effects but also provides a measuring mechanism. Nishtar (2011) observed that due to an illness of the breadwinner, people living in the developing countries suffered problems including financial loss, irreversible malnutrition, and termination of schooling, just to name a few

Heltberg and Lund (1994) reported that almost all people undergoing health-shock have faced hardship. Especially, in Pakistan, where each family has to bear around 10,400 PKR per shock. Similarly, in (Howlader, 2013), different socio-economic factors and their impacts were studied. It was reported in (Howlader, 2013) that 63.8% of health expenditure was out-of-pocket, i.e., from the pocket of the patient, which resulted in financial losses. In (Binnendik et al., 2012), health related "hardship financing" for poor households in an Indian town Orissa was studied. The authors investigated factors influencing the risk of hardship financing with the use of a logistic regression. It was observed that in rural areas, most of the households were facing financial hardships due to indirect and/or long-term costs of healthcare. In Orissa, 80% of spending on healthcare was out-of-pocket for which households either borrow money at higher interest or sell their assets.

From the various studies (Binnendik et al., 2012; Victor et al., 1976), it is quite evident that the unpredictable timing of health issues and immediate need for large funds for healthcare in addition to the distance to health facilities could increase the risk of hardship financing.

2.4 Health-Shock: Case of Pakistan

Pakistan has a projected population of approximately 191.71 million and witnesses 4.2 million births per year (GoP, 2014). It is the 6th largest country in the world in terms of population, where 39% of women have no prenatal care, and 61% of births are operated by unskilled professionals. Pakistan is ranked 6^{th} on the table of 22 tuberculosis (TB) burdened countries. 1/4th of Pakistani adults are obese, 40% of men are smokers, 30% of health facilities are not working, 19% adults are malnourished, and 30% of under-five children are malnourished. In Pakistan, the spending on doctor's fee per capita is 37.09 PKR, and 28.25 PKR is the spending on medicine supplied per capita. There are only 250 psychiatric beds for the population of 25 million in Khyber Pakhtunkhwa province. Pakistan registers 3 million cases of Hepatitis B and C. 64% of 18-44 years old have poor dental health, which increases to 93.6% for 45+ years old. The injuries and accidents account for 11% of Burden of Death (BoD) across Pakistan (Nishtar, 2011; Nishtar et al., 2013).

Pakistan is a country, where there is just one hospital bed per 1647 people, one medical facility per 11413 people, one doctor per 1099 people and one dentist per 13441 people as shown in Table 2.1. As a result, health-shock in Pakistan accounts for 54% of all types

Table 2.1: Highlights of health sector of Pakistan (GoP, 2014).

of shock, which is almost double that of economic shock at 28%. Fig. 2.2 presents the percentage of different shocks, experienced by a common household in Pakistan (Nishtar et al., 2013).

In Pakistan, only 26.62% of population is receiving the health coverage and 73.38% has to pay out-of-pocket in case of any medical-related situation as shown in Fig. 2.3 (Nishtar, 2011; Nishtar et al., 2013). Since 1992, communicable diseases like Diarrhoea, Malaria, and Hepatitis have been controlled; however, the non-communicable diseases like renal/urinary illness, nutrition disorders have increased. In 1947, at the time of independence, Pakistan

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

Figure 2.2: Different categories of shocks faced by households in Pakistan (Nishtar et al., 2013).

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

Figure 2.3: Percentage of Pakistanis receiving health coverage (Nishtar, 2011; Nishtar et al., 2013). Here "*" represents employees and their dependents whereas "**" represents population covered.

had a population of 84 million; however, with the current growth rate of 2.2%, Pakistan could easily be a nation of 350 million by 2050 (GoP, 2014).

According to Government of Khyber Pakhtunkhwa (GoP, 2010), Pakistan spends $9.3 per person on public health spending, which is far below the internationally recommended figure of $60 per person (Jafar et al., 2013; Nishtar et al., 2013). Pakistan has $8-9$ million infected people with Hepatitis C and its variants; where, 620,000 get tuberculosis and another 410,000 get TB every year—out of which 59,000 dies. Furthermore, there are half a million malaria victims reported in Pakistan every year and 5,000 die because of rabies (GoP, 2010; Jafar et al., 2013; Nishtar et al., 2013). Pakistan has 3.6 births per women, which is the highest in the region. Moreover, Pakistan has the third highest number of maternal, fetal, and child mortality deaths in the world with 57% of neonatal deaths within 72 hours of birth (GoP, 2010; Jafar et al., 2013; Nishtar et al., 2013).

In Pakistan the predominant means of financing healthcare costs are borne out-of-pocket and a majority of people face financial hardships as a result of healthcare payments (Mahmud et al., 2014). Here, healthcare costs account for more than 70% of the economic shock faced by poor households, especially in rural areas. Here, it is worth mentioning that Pakistan's healthcare model does not protect its people/patients from financial risks, which becomes one of the main reasons of health-shock (Jafar et al., 2013; Nishtar et al., 2013).

Moreover, the socio-economic, geographical, and cultural norms are affecting the health of almost 191.71 million Pakistanis—especially, of those poor people, who belong to rural and tribal areas. For instance, women in rural and tribal areas of Pakistan are not allowed by their men to consult a male doctor during pregnancy, which results in higher Infant Mortality Rates (IMR) and Maternal Mortality Rates (MMR) (WHO, 2007). Table 2.2 shows the IMR (count of the number of infants that die before their first birthday in every thousand infants (WHO, 2007)) of Pakistan in comparison to other countries (WHO, 2014). Currently, there are only 25 countries that have a higher IMR than Pakistan. Another useful measure for assessing children's health is their weight. The experts have figured out a scale that lists out the appropriate weight for healthy children at any given age. In 2001, the percentage of under-weight children, less than five years old, was 32% in Pakistan. Similarly, MMR in

Table 2.2: IMR of different countries (WHO, 2014)

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

1990 accounted for 490 deaths per 100,000 live childbirths in Pakistan for causes directly related to either pregnancy, or how it was handled (WHO, 2007). Another important factor that is directly related to IMR and MMR is "appropriate pregnancy spacing". Generally, it is recommended to maintain a 2.5 to 3 years gap between pregnancies which is vital for the health of both mother and child. However, such spacing is only possible with proper awareness, availability of the necessary equipment, and birth control (Hameed et al., 2011). According to (Hameed et al., 2011), around 27% of the Pakistani couples who preferred to use some sort of birth control did not find it available in their local region.

2.5 Summary

A brief discourse on the subject of health-shock—its causes, effects, and impact on general health of people has been provided. Here, it is worth mentioning that considering all the other types of shock (i.e. natural, economic, health, agricultural, law and order, and family matters), the percentage of the population suffering from health-shock (54%) remains the highest in Pakistan. Moreover, one of the biggest challenges in Pakistan is the availability of healthcare data. Currently in Pakistan, there is no publicly available dataset that can help to understand health-shock or the reason for its occurrence. If such a dataset was available the Iron Triangle model of healthcare (William, 1994) could be applied to give more of an understanding of health shock which would take into account the socio-economic, geographical, and cultural norms of the region.

To cater for health-shock and its consequences, there is a dire need to understand the socio-economic, cultural, and geographic norms of people living in rural and tribal areas of Pakistan. To achieve this would require a large-scale survey. The data collected from such a survey, once analyzed, could then be used by government for policy making. Such surveys and datasets can be helpful for the government, who could use this dataset and resulting analysis to form policies. Moreover, intelligent approaches can be applied on the large-scale dataset to identify patterns in the data which will allow further investigation of health shock. It is therefore essential to review the existing approaches and how they are applied in such domains.

The next chapter will present a literature review, highlighting the state-of-the-art approaches so that an appropriate approach can be selected for further enhancement and adoption.

Chapter 3

Literature Review

3.1 Introduction

A comprehensive review of the existing literature regarding healthcare, health-shock, and the involvement of Information and Communication Technology (ICT) in the contemporary healthcare system has been discussed in this chapter. The main goal of this chapter is to review the existing approaches related to health factors affecting the health and ICT based systems, importantly to investigate an artificial intelligent approach that can be used to analyze and model the data while taking into account uncertainties in the dataset.

The schema of the chapter is as follows: Section 3.2 provides an overview of the state of health—factors affecting the health and its consequences in the underdeveloped and developing countries; Section 3.3 provides a discourse regarding the intervention of ICT in healthcare; Section 3.4 narrates the story of cloud computing and its usage in healthcare applications; Section 3.5 gives an introduction to intelligent computation techniques and a comparison of existing approaches; Section 3.6 advocates the case of using fuzzy logic for health-shock prediction and why it is a better approach with qualitative parameters and subjective research; and Sections 3.7 provides a summary. Moreover, a summary of the literature review is given in Appendix A.1 - Appendix A.3.

3.2 Overall Health Scenario: Underdeveloped and Developing Countries

To understand the healthcare system of underdeveloped and developing countries, a study was conducted to investigate the impact of variables such as age, gender, religion, area of residence, income, education, body mass index, smoking, employment, marital status, health expenditure, health service quality, working conditions, and living conditions (Howlader, 2013). Here, it was observed that in Bangladesh 63.8% of health expenditure was out-of-pocket which resulted in financial losses. Furthermore, it impacts the lives of people, particularly from lower income groups. Moreover, it was observed that the quality of healthcare service (by the service provider), education, religion, marital status, wealth, and working environment tends to have a positive impact on health status. However, habitual smoking and unemployment has a negative impact on health. In the respective study (Horowitz and Savin, 2001), Probit and Tobit regression models were used to investigate the effects of socio-economic and demographic parameters on health status. Subsequently, it was discovered that the people with bad health were facing tremendous financial losses and could not easily find employment opportunities in comparison to people with good health. Generally, unpredictable timing of health issues and the immediate need for large funds for healthcare could increase the risk of financial hardship (Bernstein, 2009; Iqbal et al., 2011; Victor et al., 1976; William, 1994) whereas possession of assets and regular income-flow are predictors of lower expected financial hardship.

In (Binnendik et al., 2012), "hardship financing" for poor households in an Indian town Orissa was studied. Here, it was observed that rural households mostly were subjected to financial hardship due to the indirect and long-term costs of healthcare which resulted in the withdrawal of a child from school and/or skipping a meal. Moreover, to cover the healthcare cost, poor people borrowed money on interest or sold their assets. During the survey, the population was divided into two equal groups, namely: "member households" and "non-member households". The "Member households" were those households that included at least one person who was a member of a Self-Help Group (SHG). For most indicators, it was observed that the difference difference in healthcare utilization and cost between SHG members and non-members was not significant for most indicators, especially when it comes to expenditures for outpatient care, hospital admittances, maternity, and distance to people's choice of hospital/clinic. However, SHG members in comparison to "non-members" were more likely to suffer financial hardship.

In (Victor et al., 1976), it was observed that significant economic benefits can be achieved by improving healthcare in developing and developed countries. Thereby, the countries should work on human capital—especially, health. Generally, health can contribute to economic outcomes in high-income countries through higher productivity, higher labour supply, and improved skills which can increase the financial resources for investment in physical and intellectual capital (Suhrcke et al., 2006). Furthermore, children with better health and birth weight can be expected to attain higher educational levels; therefore, be more productive in the future, with the prospect of a longer and healthier life (Suhrcke et al., 2006). Moreover, healthier individuals, with a longer lifespan in front of them, tend to have greater incentives to invest in education and training as they can harvest the associated benefits for a longer period. At the same time, there are many costs associated with being ill as bad health affects family members as well as the performance of the individual.

Similarly, the effects of health indicators such as adult survival rates (ASR) on economic growth rates in developed and developing countries were studied (Bhargava et al., 2001). It was observed that the ratio of skilled to unskilled labour was likely to be negatively associated with total fertility rate (TFR). These factors suggest that TFR is likely to be negatively associated with economic growth whereas ASR for a country is likely to be influenced by economic development, access, quality of medical care, and the public health infrastructure.

In (Smith, 2007), health patterns of people belonging to lower "socio-economic status" (SES) were discussed. It was observed that new chronic conditions were related to household income, wealth, and education. Furthermore, dimensions of SES—income, wealth, and education in order to predict future health outcomes were also discussed. It was observed that higher-SES people may have better access to medical care, more information about appropriate medical practice, less strenuous jobs, or access to more material inputs that improve health. Moreover, they may live in more health promoting environments whereas people with lower SES do not get promoted and/or expelled from their jobs. Here, it is worth mentioning that forced occupational change can bring harm to people's health and health-related behaviour; especially, during economic transition (Lazareva, 2009). It increases the probability of smoking and alcohol consumption, which has adverse effects on health in the long term in addition to job insecurity and job loss.

Similarly, medical waste and its adverse effects on the society were studied in (Fuchs, 2010). It was reported that it contains more risk to the patient than potential benefit. However, such risks can be reduced by cost effective care which can be provided by introducing healthcare reforms and with the introduction of new technologies to healthcare.

An extensive study regarding various interventions by governments to help in providing a balanced distribution of health services in China was conducted (Liu and Tsegai, 2011). Here, participation was voluntary and it was required for households to be enrolled as a unit. Surveys showed that households with married heads, poor sanitation environment, high ratio of children and chronic members were more likely to enrol. Such interventions could help to reduce health risks by encouraging poor rural households to visit a doctor for minor diseases at early stages before they turn into serious diseases.

3.3 Healthcare and ICT

Currently, healthcare providers have a seemingly infinite combination of graphic, textual, coded, statistically analyzed, collated, filtered, non-filtered and malleable data (Lindberg et al., 2013). As a result, healthcare organizations endeavour to construct comprehensive summary views of patient's medical records. Here, the basic underlying structure is the Health Information Integration Framework (HIIF) which creates comprehensive health record views based on information from various sources.

An analysis of joint modelling of Revealed Preference (RP) survey data and qualitative choice data was studied in (Iles, 2011). Here, the two systems of medicine operating in India (Allopathic and homeopathic) were studied. The gains of using the Mixed Multinomial Logit (MMNL) and Generalized MMNL models to jointly estimate RP and qualitative choice data were analyzed. The MNL model is used for modeling healthcare provider-choice data while the Nested Logit (NL) model is also used in several healthcare demand studies. The RP and SP data comes from rural north India and relates to outpatients' choice of doctor to treat a severe fever. A total of 1,174 respondents from eight villages were sampled to obtain information on individuals including religion, age and healthcare provider visited. Price may be negatively related to the probability of seeing a government MBBS doctor. The market share for unqualified doctors, according to the RP data, was 59%; while government doctors, according to SP data, was 51%. The results here show that a combination of a balanced RPSP dataset, with a mixed MNL error component model, can be useful.

Furthermore, an exhaustive survey was conducted for the use of ICT in healthcare (Lindberg et al., 2013). This research has provided definitions of tele-health and tele-medicine along with e-health in addition to a proper appraisal mechanism to weigh the ICT interventions in health. However, it was too generic to conclude any direct ICT based interventions for the improvement of health systems in a country such as Pakistan. Similarly, a summarized report on HIS-EVALs was presented in (Ammenwerth et al., 2004). It provides a historical account of work in this area and highlights the importance of different disciplines for improvement in health information systems such as bio-statistics, medical informatics, psychology, computer science, and health economics. Furthermore, Pakistan‘s health system was defined at the district, provincial and federal levels in addition to the key issues of the poor and declining health landscape of Pakistan.

In order to reduce health expenditures by adopting health information technology (HIT), a large study was conducted in U.S. (Callaway and Ghosal, 2012). HIT allows clinics and physicians to manage information. Furthermore, it provides a secure exchange between healthcare consumers and providers to reduce medical errors. Here, various clinic-level characteristics were combined with geographic location-specific information to create a comprehensive dataset for examining factors that influence HIT adoption decisions. The driving force behind technology adoption was found to be related to characteristics such as "firm size". Each firm had a different level of benefits and costs that would result from adopting the new technology. Furthermore, it was observed that the cost of adoption usually decreased over time, while the quality improved.

Similarly, if the gap between the clinics' costs of operating with both old and new technologies increases, the number of clinics adopting HIT will increase. Studies show that the clinics that see a patient more frequently and for longer periods of time may want to adopt HIT for better record-keeping, while an Urgent Care clinic may not. Some characteristics such as physician rate, income, educational status and wages of medical assistants, percent of smokers, percent of obese people and percent of population without health insurance were also studied. It was found that the hospitals in states with more stringent privacy laws were less likely to adopt HIT. The location of a clinic had a strong effect on the probability that it adopted HIT, while median income had a small positive impact.

3.4 Cloud Computing and Big Data Analytics

Cloud computing can be described as combination of computing resources that can be mutually shared and configured within the work domain to provide ubiquitous computing architecture (National Institute of Standards and Technology). Cloud computing provides on demand and measured services with a flexibility to utilize cloud resources such as network access, as and when required (Mell and Grance, 2009). In general, the concept of cloud computing has been employed in many fields in addition to the healthcare domain. In healthcare, cloud computing is widely used in information sharing, data analytics, data portability and ubiquitous access to healthcare facilities.

Due to scalability and flexibility in healthcare provision, cloud computing has become the prominent option in global health systems. The domain of healthcare systems is diverse and highly complex comprising of a number of collaborating entities such as hospitals, pharmacies, laboratories, clinics, insurance companies and many other research establishments (Wan et al., 2010). These entities work within some standard governmental regulations. For effective working of this healthcare ecosystem, a fast, secure and efficient information sharing mechanism attains a core place. Here, health data integrity is a main challenge and a risk for healthcare organizations moving towards cloud. Cloud provides innovative technologies and tools to deploy state of the art systems specifically regarding healthcare infrastructure. Furthermore, effective collaboration among these entities can unleash the potential of cloud computing to provide health facilities from cities to whole countries.

Furthermore, with a rapid increase in data sharing, privacy of medical records has become an important issue for healthcare providers. As the patient's data is highly sensitive, there is a certain set of rules established by government entities, such as the health insurance portability and accountability act (HIPPA), to avoid any privacy violation. At this level, a high cost is required to build such secure systems which might impact the performance of healthcare business models with respect to quality and access. However, cloud computing is an emerging option to improve the access to, and running of healthcare systems without decreasing quality. It can make health information available anywhere any time by moving health data to cloud storage. Furthermore, to contain budgeting and integration costs, cloud computing provides a state-of-the-art platform for customers, ready to use, without compromising the quality of care.

Moreover, to enhance the quality of healthcare delivery at an optimal cost, it has provided opportunities to patients as well as other healthcare entities such as clinics, medical stores, insurance providers, laboratories and hospitals. Due to the high cost of infrastructure, many organizations prefer to build and deploy their applications on cloud to meet their system specific needs. Moreover, cloud provides a variety of choices in terms of platform resources such as multiple operating environments, test beds, deployment servers and database options. Furthermore, cloud provides scalability in terms of load as the infrastructure is already built to use by customers which ensures high availability in case of increasing network traffic. Thereby, organizations are able to achieve their goals on a very nominal cost as compared to building their own set ups.

Cloud computing provides interoperability within healthcare systems and will have a growing future ahead. Significant research is in progress to make cloud experience more worthwhile. Right now, researchers have identified three types of cloud environments, given as, public cloud, private cloud and hybrid cloud. They provide a varying number and types of cloud services in terms of data sharing, management and processing. E-Health applications mostly use public clouds (e.g. Google app engine) and their practical implementation and testing has been demonstrated by different studies (Wu et al., 2012b). Another example of cloud is Windows Azure (Benharref and Serhani, 2014; Lu et al., 2014) and Amazon EC2 (Bhange and Hiray, 2012; Sabah et al., 2011; Wooten et al., 2012).

Furthermore, some of the organizations are leveraging their needs by establishing their own private clouds to replace legacy infrastructure. Private clouds ensure security of their data and give more control over deployed environment, i.e., servers, security firewalls, band width with a transactional approach to avoid risk of data breach. Similarly, the cloud community model is another approach to share data among different healthcare entities. It provides a shared set of resources regulated under specific security needs and data sharing policies. It also provides an interconnected platform to healthcare providers and health organizations to overcome security risks. In this way, data, health records and image scans will be available to be used for diagnosis and counselling purposes within community health entities. That will aid to implement a secure and quick mode of information exchange at a lower cost. Here it is worth mentioning that the use of cloud computing will become a norm in healthcare domain in the coming years due to its potential fruitful outcomes in healthcare. Furthermore, despite the privacy and data security challenges, healthcare systems and organizations (especially in under-developed countries), can take advantage of cloud computing for the efficient delivery of healthcare services in remote and medically deprived areas around the globe.

3.4.1 Factors to be Considered Before Moving to Cloud

Deployment of clinical information systems to cloud seems promising and has associated criteria for selection. Various studies below present factors to be considered while moving to cloud. For instance, Low and Chen (2012) suggested that cloud service providers should be carefully evaluated before selection, as the commercial vendors are dramatically increasing their services for consumers with state of the art infrastructure on multi option payment plans. Further to this, the authors proposed a provider selection evaluation model based on the fuzzy Delphi method (FDM) together with the fuzzy analytical hierarchy process (FAHP). Here, they suggested the use of evaluation criteria based on certain decision factors such as reliability, system usefulness, service quality, professionalism of service provider and system usefulness. Based on these factors, selection and evaluation of specific service providers becomes less complex and more robust from an investment perspectives as well. Contrarily, Yoo et al. (2012) have chosen a rather more conventional approach of setting up a private cloud. Here, based on a virtualisation approach, fixed virtual desktops consisting of around 400 virtual machines were used to share information throughout the hospital, following an easy approach to access the hospital information management system (HIMS) from all devices within that hospital. After that, they performed a cost-benefit analysis covering five years of cost estimates and identified that this approach would reach break-even point in the fourth year of investment.

The main objective to move to cloud computing is to facilitate collaboration between healthcare workers, health units, remote locations, and to achieve ubiquity in delivery of healthcare services while reducing the high costs of resources to an affordable level. The growing amounts of electronic data and scanned digital images require a vast amount of storage space and high availability which can only be provided by cloud services in this modern era. Here, it is worth mentioning that there are certain needs that need to be satisfied by a system before organizations make the decision to adopt a cloud computing solution. A brief discourse regarding this is given below.

3.4.1.1 Architecture of the Healthcare Systems

It is worth considering the nature of the healthcare system whether its architecture centralized or it is distributed to multiple locations. Here, adopting cloud platform would surely help to ease communication and collaboration among remote geographically distributed healthcare delivery units.

3.4.1.2 Deployment Cost

Another factor to consider is, whether moving to cloud would reduce the cost of services being provided by the system. Moreover, by developing cloud based systems, the number of field workers and facility staff providing monitoring services in remote units will also reduce. In this way the associated management costs will reduce accordingly.

3.4.1.3 Data Sharing Across Systems

Previously, in most of the electronic medical records (EMRs), data was stored in the databases of healthcare providers lacking mutual information exchange and collaboration (Zhang and Liu, 2010). But, nowadays, all of the modern healthcare systems are able to share and expand electronic data among other systems. Here, cloud computing provides extensive support for electronic data sharing of doctor prescriptions, diagnostic tests, insurance information, and medical images across different systems in a secure way. Moreover, due to frequent travelling for business and personal reasons there is a need to have medical history available anytime, anywhere in the world to avoid any serious risk. For instance, Microsoft's HealthVault (Microsoft, 2015), is a cloud based healthcare system for sharing image scans and fax records between doctors, which afterwards converts them to digital formats. HealthVault has multiple features including the means to connect various health service providers e.g. labs, pharmacies and clinics to a centralized place to store and share health records. Users have rights to add health related data from various fitness and health devices to their health records, by using the HealthVault connection center. Furthermore, there are number of other options besides HealthVault such as Oracle's Exalogic Elastic Cloud and Amazon

Web Services to store personal health information on cloud (Kuo, 2011). This collaboration is a key strength of the cloud computing platform and compliance to information sharing format according to HIPPA laws and HL7 standards retains a core importance. Furthermore, putting health information on cloud will not only ensure more treatment options available across the world, but will also increase the patient satisfaction level for healthcare services provided with less possibility for errors.

3.4.1.4 Reliability of Legacy Systems

Generally, moving to a cloud platform needs a very optimal cost for a wide variety of provided services on public cloud. However, the functional reliability of legacy systems is a key factor to consider while upgrading to cloud. For instance, legacy systems might have limited network handling capacity for geographically distributed locations. Here, when upgraded to a cloud platform, the system might draw a heavy traffic flow on the network from time to time, causing latency and a low quality communication experience. Therefore, reliability of the existing system is of extreme importance when moving to cloud. Furthermore, to ensure uninterrupted ubiquitous access it is necessary for healthcare providers to maintain a high availability of services with the minimum amount of downtime. Here, one approach to achieve this is to make clusters with multiple nodes (Kupferman et al., 2009). Moreover, cloud can be programmatically controlled and monitored to provide a smooth and scalable access during off-peak and peak hours. Hence, it will dramatically decrease maintenance costs by monitoring consistent uptimes of cloud healthcare services. In addition, open access to a large audience will also give motivation to service providers to consider security and privacy issues while moving towards cloud platform (Zhang and Liu, 2010).

3.4.1.5 Privacy and Security Risks

In spite of all the benefits such as information sharing, data portability and high availability, there are multiple personal data security risks associated with cloud computing. Healthcare providers ensure the availability of data while enforcing HIPPA privacy and security compliance. It is the duty of cloud service providers to keep intact the patient data integrity which is at the heart of cloud computing in healthcare. Furthermore, availability of data could be a target of malicious attacks causing data loss or data breach due to insecure API implementations. Hence, guaranteeing complete data privacy is the most sensitive task regarding the future of cloud computing in healthcare. For this, private clouds are implemented initially to bottom-line the priorities regarding healthcare organizations (Wan et al., 2010). It might prove to be a good approach to see which of the aspects of cloud computing are beneficial for the organization in the first place, then making effective use of them. Here, it is worth mentioning that healthcare costs, quality of service, data security and disaster recovery are considered to be the top and foremost priorities of the current era (Wan et al., 2010). Furthermore, changes and improvements in the cloud infrastructure can meet most of the user priorities mentioned previously. Another study (Regola and Chawla, 2013) illustrates that, these changes and improvements should be of prime importance when moving health data to a commercial cloud maintained by third party cloud hosting, due to security concerns. Furthermore, the data security aspect in cloud was also studied by authors in (Rodrigues et al., 2013). Here, the researchers identified the security risks associated with storing electronic health records (EHRs) on cloud servers. The authors presented a comprehensive review about data security and privacy concerns with vendors providing cloud services. They suggested that an external standard organization should perform audit of the cloud providers before deploying healthcare systems to their servers.

Work in (Regola and Chawla, 2013), emphasized the need for HIPPA compliance for health data sharing among healthcare entities using cloud services. Authors portrayed the benefits of cloud computing in terms of security, access, and cost effectiveness. Afterwards, they provided a brief set of information specifying the impact of the enforcing HIPPA laws. Furthermore, by enforcing security compliance researchers would be able to implement more robust approaches to improve healthcare delivery using cloud technology. Similarly Chard et al. (2011) examined an approach to use cloud services to achieve high scalability and HIPAA-compliant data security within healthcare systems. Here, the authors proposed a cloud-based SaaS NLP prototype for several hospitals. The main aim of this prototype was to provide the ability to perform various operations on medical data such as extraction, procession, management, and comparison of healthcare data. Similar NLP service was given by (Christoph et al., 2014), here the medical texts were de-identified before they were pushed to cloud for storage. Moreover, patient information was encrypted as a security parameter for patient data protection.

Extending further, authors in (Rodrigues et al., 2013) mentioned that “cloud-based EHR must maintain the same level of data security as data stored in the servers of the health care provider”. However, various approaches are presented in the section to follow regarding security of healthcare data.

3.4.1.6 Privacy Control Mechanisms

Among the major concerns about cloud services are basically uptime and performance of the cloud platform with a guarantee of data security (Hussain et al., 2014). The literature shows various approaches to ensure data protection and security in cloud computing. One of the common techniques to protect data effectively is access control. For instance, in (Liu et al., 2013a), an identity based encryption (IBE) is implemented to provide access control to patient health records (PHRs) and this cryptography mechanism simplifies the key management. Another mechanism is attribute based encryption (ABE), which is the most preferred encryption in healthcare cloud computing. In (Onik et al., 2012), cipher text-policy ABE is used to make the security manager module act like a real time administrative person. Similarly, Barua et al. (2011) and Narayan et al. (2010) propose another type of access control to give patient-centric ABE. Similarly, one of the encryption strategies is the cloud-based ABE which was implemented for emergency mobile access to PHRs. Likewise, to detect normal and emergency access to data, a mixed usage of ABE and IBE schemes is proposed by other authors (Huang et al., 2012; Lounis et al., 2012; Tong et al., 2013) for a more robust data encryption. Furthermore, based on ABE, there is an approach called role based access control which provides role based authenticated access to corresponding operations within a system. In the same context, Liu et al. (2011) discusses the security of various features of healthcare data such as authorized access, controllability and data traceability in terms of cloud-based privacy-aware role based access. Sharma et al. (2012) implemented a task based control, to check if the access has been granted previously to the healthcare cloud.

In addition to access control, there are many other security protection techniques such as secret-sharing schemes (Ermakova and Fabian, 2013), secure index implementation (Alabdulatif et al., 2013), trusted virtual domains (Löhr et al., 2010) and watermarking method (Yu et al., 2012). These mechanisms were proposed to ensure security and privacy of health data in the emerging field of cloud computing.

Forensic methods are also considered as the best approaches to minimize the risk of data breach at the cloud owner-end of e-Health cloud architecture (Ahmed and Raja, 2010). Here, the authors further elaborated that forensic security methods are much more effective in order to ensure data privacy, information confidentiality and tracking of activities in a cloud environment. Further to this, the authors proposed computer forensic tools (CFT) as a security model for healthcare data cloud. With the help of data loggers and sniffers installed on the cloud machines any volatile information can be easily captured. Contrarily, certain authentication, authorisation and access control mechanisms are used to protect other non-volatile information. The mechanisms also suggest the need for a periodic certification by a third party audit (TPA) and the supporting of the e-health cloud with standard legal considerations. To automatically keep check on activities of the cloud owners, TPA observes the forensic and anti-forensic tools periodically. Here, the main challenge for e-Health Cloud with CFT is that it does not physically prevent the volatile actions even though it has all the potential required to ensure privacy.

Löhr et al. (2010) emphasize security issues related to end-user platforms and third party storage platforms for healthcare data. Here, the authors argue that end users make use of cloud platform for multiple applications in addition to merely accessing healthcare data. Owing to the need for access to multi-application users, cloud platform becomes more malware prone and vulnerable to suspicious intrusions which can steal the user's credentials and other private data. This adds to other outstanding issues of data storage and transfer in cloud environment. Here, one approach suggested by authors is to design trusted privacy domains (TVD) as a solution for issues regarding the privacy of medical data of patients. According to the suggested solution, the system will divide the execution environment for individual applications within isolated domains thus authorized users will only be able to access medical data within corresponding TVD. Moreover, data encryption can be performed automatically while storing data to external storage by using a TVD specific authorization key. Here, it is worth mentioning that key management and policy implementations are totally transparent and open to TVD users. Whenever a client requests to join a TVD, the cloud infrastructure dynamically performs verification and client authorisations by enforcing policies. Further to this, the authors described the policy enforcement procedure by security kernel without user interaction. Here, the connection among different platforms is secured by using a virtual private network (VPN). Furthermore, authors described the case that whenever data is exported to external systems which are not associated with any TVD, a security gateway is specifically introduced to monitor data export activity. Here, it is worth highlighting the main challenge faced by this approach which is having a security module installed such as the trusted platform module (TPM) on devices, which cannot be installed to all of the patient devices. Furthermore, in this type of cloud, scalability becomes an issue because the privacy domain was implemented on the platforms used by the health services providers, for each application.

Moreover, a privacy framework was introduced by authors in (Kaletsch and Sunyaev, 2011) to support the secure exchange of medical data. Here, authors suggested the need to use an online referral and appointment planner (ORAP) to securely transfer information to healthcare specialists. The ORAP security model, considered the practices space of physicians to be the only trusted environments for EHRs where data encryption was completed before leaving the space and decryption was performed before the physicians received the data. According to the implementation of this framework, even the patient's home was not considered as secure thus EHRs access was restricted for patients at home as well. The main drawback of this system was that it only provided the feature of transferring EHRs data without allowing any patient-centric operations such as collaboration tools, action plans, view/update entries and medical history of patients.

So far the discussion about various approaches gave an insight to patient centric privacy solutions within cloud enabled systems. In the presented solutions, access control and data encryption requires authorisation certificates, encryption keys and the implementation of policies and privileges. These approaches provide more control for ensuring patient satisfaction in the use of healthcare systems. However, this fact has made patients more responsible about with whom, or how much, information they are sharing. Patients need to be more serious and aware while disseminating their medical records. Furthermore, in case of emergency, the patient centric approaches may cause trouble if the patient is either not present to grant access or incapacitated through illness or accident. In these scenarios, security measures will incur an increased overhead for healthcare organizations in order to initiate proper treatment processes on time. Moreover, handling multiple service providers with a number of patients would cause more complexity in the deploying of policies for each patient and service provider. Therefore, to solve this complex situation, a more flexible solution for applying healthcare security policies is needed.

3.4.2 Applications of Cloud Computing in Healthcare

Wang et al. (2011) argued that by employing a model which is cost-effective from the point of view of both the patient and service provider the quality of service for cloud solutions could be improved and maximum benefit achieved. Also, a better healthcared system requires coordination of medical process with a decrease in corresponding maintenance and infrastructure costs. Cloud computing fulfills the quality, access and corresponding cost needs effectively.

Furthermore, cloud computing has emerged as a modern business paradigm for many industries including biomedical information sharing (Rosenthal et al., 2010). Authors described the possibility of improving health services with cloud computing (Kuo, 2011) and presented the future opportunities and relevant challenges in the implementation of healthcare clouds (Kuo, 2011). To provide healthcare service to deprived and underdeveloped areas, Subrahmanya et al. (2010) proposed a solution using cloud computing. The article lacked any illustration or explanation of the approach that the author proposed for further evaluation. Moreover, it is not mentioned whether the solution provided would be a cloud application or a simple telemedicine service for under-developed areas.

Oshidori-Net2 is a server-based computing healthcare system designed for six Japanese hospitals, which later on authors recognized as cloud based technology. Authors further described that virtual servers and virtual routers were used in this architecture, but no further explicit details about its cloud deployment was given (Kondoh et al., 2012). Similarly, Shih et al. (2012) and Liu et al. (2013a) conducted a study to ask questions from healthcare professionals and technology experts about the pros and cons of e-health document sharing among various institutes on a web platform. In this study, 15 e-health technology experts and around 65 organ transplant healthcare professionals from China and Taiwan were involved. However, the authors did not give any explicit details of the cloud services used for secure document sharing in the system. Furthermore, Rajkumar and Iyengar (2013), presented the concept of a peer to peer network to transfer medical records and patient histories between different health entities such as hospitals and ambulances. Here, each cloud connected with the community hospital provided data storage to share patient data with ambulance nursing staff. Furthermore, the author described the cloud application and overall architecture of the test bed with the possible benefit of reducing the death rate in emergency care resulting from delayed transportation when there was limited patient information available to the field nurse.

A similar concept, named as "cloud emergency medical services", was given by Koufi et al. (2010). They applied the concept of platform-as-a-service and software-as-a-service to design the system. The authors further proceed to depict details about the cloud implementation of the system prototype on a laboratory cloud environment. Here, the data was stored on multiple nodes within the cloud.

Cloud cardiology was another system to share electrocardiogram (ECG) report simultaneously inside or outside of the hospital (Fujita et al., 2012). Authors, discussed that the system used a cloud server, however, no details were provided about the implementation or specification of the server being used. Furthermore, a study for ECG transmission prior to hospitalization was carried out (Al-Zaiti et al., 2013). Here, an overview of standardize protocols were studied which were used by different vendors. Moreover, cloud computing proved to cut down the initial investment cost for adopting this technology. Utilizing the similar concept, Fong and Chung (2013) presented a mobile cloud-based healthcare service for non-contact ECG monitoring within a healthcare unit. However, the article further portrays that the software used a client-server architecture without mentioning cloud implementation. Furthermore, Wang et al. (2014) presented a conceptual model of a hybrid cloud computing environment for data storage, by wearing personal health sensors (i.e. like ECG sensors) to perform processing tasks. Here, the cloud servers were used to perform an intensive processing task rather than on mobile devices in order to save battery life.

Work regarding a telemedicine ECG service based on cloud computing was also reviewed in detail by Hsieh and Hsu (2012). The authors presented a complete description of processing, visualization and management of e-Learning services that were deployed on the Microsoft Azure cloud environment. Further to this, they described the factors and costing of using this cloud platform in terms of CPU hours and GB storage being used during each month. Similarly, development of e-Science central—an application utilizing cloud as platform as a service is presented by Hiden et al. (2013). They provided a comprehensive review on different aspects of data security, enforcing workflows and service execution with data storage services for e-Science central. Here, they studied three case studies, out of which one was the MOVEeCloud project—a pilot investigation for specialists to gain important insights about the physical activities of patients using data uploaded from wearable accelerometers to the e-Science cloud.

The monitoring of patients during post-hospitalization days is of immense importance to avoid re-admissions due to any carelessness. Therefore, a large scale improvement is required in monitoring systems, specifically for elderly people. Following the concept, caREMOTE—a pilot prototype developed for cancer patients, was implemented to monitor and report patient situations to telemedicine systems via mobile devices (Cheng et al., 2011). Here, the cloud services were built on Google app engine (GAE) and a big data table was used to store health data. Further to this, authors discuss the usage of GAEs sandbox technology and its advantages in isolating the caRemote database in cloud because of security concerns about the medical data. This isolation protects patient data from being stolen or violated. Moreover, to provide a more robust patient data security, authors described the implementation of advanced encryption standard (AES) in future releases.

Furthermore, the concept of remote monitoring and reporting services was also studied for the neuroscience field (Zao et al., 2014). In this paper, authors used an online EEG-BCI system, which was a brain computer interface (BCI) system having mobile devices and wireless headsets to predict cognitive states according to real life situations. Here, the study included health issues for both patients and individuals from daily life. Further to this, they proposed a prediction model for cognitive states to find certain matching segments in data related to personal, geographical and other specific situations. Cloud servers were proposed to deliver the required services to conduct semantic searches within the corresponding data segments. A similar approach was implemented to design a system that uses a smartphone sensory data to identify routine and ultimate life style patterns (Hussain et al., 2014). The authors also presented a predictive analysis environment and its architecture. They used Hadoop, implemented on four host machines, for their "big data" technology due to its high scalability on a cloud platform. Furthermore, the cloud based real time health monitoring system (CHMS) was another example of patient activity monitoring. It established a cloud to use multi hop sensor networks. Here, the main focus of the system was to increase the quality of service by using networked routers and computers to ensure effective exchange of messages (Almashaqbeh et al., 2014).

Cloud computing was applied to the exchange of medical images to provide immediate information to experts around the globe. Kagadis et al. (2013), presented a theoretical review of aspects of cloud computing regarding medical imaging. Later, they talked about major application areas for cloud computing in image processing, sharing and archiving discussing the security risks involved. It was more of a visionary paper and did not provide any practical implementation of the said applications. Further to this, in this domain, studies also showed work in developing medical informatics and decision making systems. Another study explained the cloud-based picture archiving and communication systems (PACS) to store medical images (Silvaand et al., 2012) and had the potential to provide a highly flexible service called "radiology round-the-clock" (Kharat et al., 2012).

To exchange medical images in a secure way, a digital imaging and communications in medicine (DICOM) has been introduced. A DICOM compliant bridge is designed for securely sharing DICOM services among medical entities that supports the efficient provision of imaging services among different institutions (Silvaand et al., 2012). Here, the communication was performed in a secure way to eliminate privacy risks. Moreover, Sofka et al. (2012) developed an efficient system which proposed the exchange of large medical images between PACS and analysis servers to overcome transmission delays. Doukas et al. (2010) developed an Android client application to receive patient images, demographics, prescriptions and history details from a central server. Further to this, the author explained that the server ran an Amazon virtual machine and later analyzed the images by using 3G and WLAN. Here, the architecture does not give any information regarding safety measures used for medical data protection nor gives any description of other cloud features for this system.

Cloud computing enables to carry out computation intensive tasks due to its distributed processing nature, scalable architecture and ubiquitous accessibility (Langer, 2011). Furthermore, the accuracy and efficacy of the cloud computing approach is also shown in arthroplasty—a technique used to repair a damaged joint (Maratt et al., 2012). Here, the process involves templating as a part of the planning performed before a complete hip arthroplasty. Afterwards, the author discussed the results of the study commenting that the digital templating was quite beneficial and more accurate than traditional methods. Further to this, the main focus of this study was to consider the medical outcomes resulting from employing digital services in healthcare.

In (Yoshida et al., 2012), authors described an implementation of a framework for distributed image processing with corresponding performance on deployed processing units. However, the implementation used multi-core CPUs in a single machine and transmission of processed data to a cloud platform. Furthermore, CometCloud according to another study, has a hybrid cloud-grid distribution framework. The basic theme of this framework was to evaluate five out of the major methods used in image texture analysis (Qi et al., 2012). Here, the evaluation was performed on a local grid like cluster.

To focus on under-served patients suffering from diabetes (Piette et al., 2011) and hypertension (Piette et al., 2012), automated calls and information delivery methods had been proposed. This technique helped patients to improve self-management against diseases. Even though authors actually did not know about the cloud computing mechanism, they still considered using a cloud platform for their services. Moreover, Takeuchi et al. (2012) proposed a prototype of a system on a cloud platform using a mobile device to collect healthcare data about diets. Further to this, the authors claimed to implement data mining techniques to extract information for each individual. The authors also suggested the need to add a feature for dietitians to add comments about users into the system. No mechanism for accessing data from the cloud was provided by the authors. In addition, cloud computing is considered as a component of a service to develop a private healthcare cloud which should provide instant warning systems to enable citizens to manage their own health (Agu et al., 2013). Siddiqui et al. (2014), described a tele-care medical information (TMIS) system which had multiple services for patients and medical professionals to provide for the monitoring of physiological signals. A smartphone equipped with authentication possibilities was used to connect to TMIS. Here, authenticated access to TMIS protected data and the enforcement of standard privacy laws was performed. Furthermore, MyPHRMachines was a prototype discussed by Gorp and Comuzzi (2014), where application software was used to view and analyze patient data in PHR. Here, data was stored in the cloud and after uploading data to

MyPHRMachines, patients were able to access this data again from remote virtual machines that contained specific analysis and visualization. There was no need for verbal communication as the patients were able to share the virtual machine session with the corresponding care givers.

Further studies are focused on specific user groups having various health related disorders. For instance, to monitor the progress of Post Traumatic Stress Disorder (PTSD) during the treatment, an automated cloud-enabled stress disorder monitoring screen was created (Xu et al., 2014). It enabled patients to visualize and measure their stress recovery states. Here, authors suggested that tele-PTSD Monitor (TPM) should be accessible via the internet by utilizing the power of the Amazon Elastic Compute Cloud. Further details are not present about the implementation and outcomes of deployment to the Amazon cloud in this study. Similarly, another author introduced and described Intelligent Aging-in-place Home care Web Services (IAServ) (Su and Chiang, 2013). IAServ was an electronic platform for elderly people to provide them with healthcare facilities while residing at home instead of at a hospital or clinic. Later on, the authors provided brief details about the architecture of the application which included an agent base environment and knowledge layer with cloud computing services. Clearer discussion about the implementation of cloud computing was missing in this literature. Similarly, Tseng and Wu (2014) focused on ensuring a healthy life style for elderly people by using their prototype "iFit". It was basically a platform for increasing awareness about physical fitness in elderly people. iFit provided game like tasks to improve physical movements of elderly users giving expert fitness diagnosis stats and corresponding fitness levels. An expert cloud was used to exchange fitness data through web services. Moreover, iFit was able to give fitness suggestions for users to improve their daily routine. In the same way, researches conducted a study on large populations by conceptualizing the use of virtual private clouds and showed public health reporting on cloud (Jalali et al., 2012).

Work by Price et al. (2010), shows a reduction in epidemics analysis time by using services of cloud structures. In another work, Eriksson et al. (2011), pandemic influenza outbreaks were studied with associated factors. Furthermore, research by Ahnn and Potkonjak (2013), presented the theoretical explanation for creating a cloud based mobile health platform by focusing on energy efficiency. Work by Vida et al. (2012) reveals the environment of two hospital departments in Romania. Both hospitals had different clinical subsystems which were able to exchange clinical information according to HL7 standards. Further to this, authors elaborated on the architecture and cloud service deployment models, however, reasons to use cloud in the given scenario was lacking information. Furthermore, a step was also taken by the Malaysian government to use electronic HIMS. Here, the objective was to give a paradigm shift towards cloud computing for a cost effective healthcare system (Ratnam et al., 2014). Moreover, in this initiative, Microsoft windows azure was used as a part of cloud architecture. Similarly, to exchange resources between large scale general hospitals and grass root healthcare institutes, China also introduced a community cloud based medical service delivery framework (CMSDF). This system also incorporated rural as well as other smaller community health service centers (Yao et al., 2014). Further, the architecture of the framework was given where the main hospital managed the cloud based virtual desktop infrastructure, which was shared with smaller health units as software as a service (SaaS). All of the medical information was stored on cloud and was readily available to all connected grass root health units.

A brief review of a clinical decision support system (CDS) is given by Dixon et al. (2014). The authors detailed features of an application that contained patient data and sent it to a cloud server to perform analysis based on certain set of rules. Furthermore, an overview of local assessments and remotely produced results were analyzed by Dixon et al. (2014).

In general, the processing power of cloud computing is the core feature that allows tedious tasks to be performed quite effectively. There is no upfront investment for the user as the pay as per use model is utilised, which makes it a more reasonable choice for small healthcare organizations. Furthermore, studies also show how gracefully cloud computing reduces execution time by providing faster processing speeds. For instance, to simulate a clinical linear accelerator, the author (Poole et al., 2012) used the Amazon cloud and noted that when around 20 worker instances were used, a $\frac{1}{n}$ reduction of computing time was achieved. Similarly, another study (Miras et al., 2013) depicts an application of Microsoft azure as cloud platform. Authors observed that up to 64 virtual machines were used, each one of different sizes, to measure a speed-up of upto 37x. Even a more complex calculation was performed in (Na et al., 2013). Here, a cloud platform was provided by Amazon which processed up to 100 achieving an observed speedup of 10-14x. Here, it is worth mentioning that, the difference in the speedup is due to the ability of parallel execution of the algorithms along with the corresponding overhead of data communication and worker management. Furthermore, in terms of performance cost for above three examples, it was noted that the cost of using cloud infrastructure was less or equal to the cost of using an equivalent local hardware cluster. Another study (Parsons et al., 2014) provides a description of simulations of radiation dose by using Monte Carlo simulation on an Amazon cloud-based model. Here, a web application called VirtuaLinac was used for the modelling of radiation treatment components.

Moreover, researchers are proposing to share data mining models and their corresponding results between linked hospitals using cloud servers (Shen et al., 2012). Further, researchers are trying to find ways to normalize medical data for exchanging it among various health entities (Rea et al., 2012). Authors in the previous study proposed a cloud based architecture for data mining and data normalization to gain the full advantage of the power of cloud infrastructure. The authors in (Doukas and Maglogiannis, 2011) have developed a textile wearable platform based on software and open hardware. The wearable basically collects motion and heartbeat data and wirelessly stores it to cloud. This data is then used for monitoring and processing purposes. Here, the textile platform contained sensors to collect bio-signals such as pulse rate, heart rate, temperature and blood pressure. Later on, this motion data was sent to the smartphone of patients via a Bluetooth connection. After receiving the data, the smartphone forwarded it to the Google cloud services center by using rest web services. After reaching cloud storage, data was processed by using a java based application which sent health alerts to patients. Here, Google cloud charts were used for data visualization purposes. Authors provided an invaluable insight into cloud architecture by giving thorough implementation details in this study.

Ratnam and Dominic (2012) highlight the example of implementing cloud services to connect various medical institutes providing a data sharing feature between them. Here, the author proposed to use medical devices as a part of cloud architecture. A local health information management system was hosted in a cloud platform with adequate cloud resources to improve the medical workflows relevant to patient health. The authors, further depicted that different medical devices could be connected together using cloud services to perform effectively with increased processing capability.

Pirahandeh and Kim (2012) suggested an innovative approach to develop a social network with a capability to communicate between clinics, doctors, patients and other healthcare personnel. It employed a private cloud which connected all of the entities to the medical resources required. Here, two algorithms were proposed to improve patient treatment, information sharing and resource management. The first algorithm initialized the list of working groups for the proposed P2P network. This network was integrated later with the Swedish national projects network (NPO). The second algorithm was used to search a social network using selected criteria including nearest resource location and distance from the enquirer. Here, patients were able to easily search for and choose an appropriate doctor. Further to this, the authors provided an overview of the implementation.

Furthermore, huge data storage capacity is one of the most prominent advantages of the cloud platform. Various researches have implemented and proposed data sharing frameworks in healthcare domain, more specifically in e-Health. Among the pioneer studies, Rolim et al. (2010) developed a sensor based data collection framework with cloud services. Here, the sensors were attached with medical equipments and the collected data was stored in the cloud. The authentication of requesting users and medical staff was the core feature of this system. A few studies (Gul et al., 2012; Hendrick et al., 2013; Patra et al., 2012) proposed an e-Health framework for large populations at national level based on cloud models. Other authors (Patra et al., 2012), argued that their proposed solution would be a cost effective approach for dealing with patients in rural areas. Here, the authors suggested it was necessary to motivate the population of rural areas to upload their personal health records to the e-Health cloud for better diagnosis and medical facilities. This way, caregivers would be able to provide more effective suggestions to deal remotely with the complicated issues of local residents. Moreover, the integrity of their personal health data will also be ensured in the cloud environment.

Saldarriaga et al. (2013) developed a mobile application based on Android and iOS (iPhone Operating System) for ambulatory ECG monitoring. The main features of this application included a guide about the efficient usage of smartphones and the management of routine activities. Here, to acquire ECG data from patients, a Kardia board was used, then the collected data was sent to a smartphone. Further, the application was able to analyze, process and upload data to a cloud server for further processing. The application was of great importance for patients and doctors to review and monitor health status. Doctors were able to receive an ECG waveform in order to help to diagnose various health problems. The only drawback of this application was that it could work only on Android or iOS phones while neglecting all other vendor platforms. However, the work could be extended to all available platforms as per user needs.

Wooten et al. (2012) discussed an application to provide support for patient-to-patient information sharing. The work by Benharref and Serhani (2014) proposed a patient controlled information sharing mechanism using mobile devices for elderly people. Here, the patient decides with whom this information and medical data should be shared. Similarly, other authors (Sabah et al., 2011) proposed and implemented health cloud exchange (HCX) system to share medical records among services and consumers with addition privacy controls.

Work in (Gorp and Comuzzi, 2014) employed virtualisation approaches to facilitate patients by giving flexibility to build their own PHRs which can then be shared with other healthcare personnel. Here, the authorization of users and patient privacy was also maintained as the author described. Wu et al. (2012a) proposed a broker based access control mechanism for EHR's data schema. A cost effective cloud platform was established by HP, named as Fusion (Basu et al., 2012). Here, the authors described that the motivation for the platform was to share and manage large amounts of healthcare data securely. To provide a more secure healthcare billing system (Rodriguez-Martinez et al., 2012) and other law systems. (Wu et al., 2012a) suggested the concept to integrate a private cloud with EHR systems whereas other authors had proposed a hybrid cloud using a shared EHRs system (Chen et al., 2012; Gul et al., 2012). Furthermore, a mixed approach of using both a private and public health cloud for saving the medical data of patients, has also been identified (Chen et al., 2012). The purpose of this approach was to give authority to the owner of the data to decide whether to store this data on a public cloud or to keep it on a private local cloud. In another approach (Dixon et al., 2013), the authors implemented an exchange system for medical data between two remote located service providers on a community cloud platform. The ultimate usage of this system was for chronic disease healthcare.

Many studies still question about the maturity of cloud computing services that different vendors are providing, which is why healthcare organizations are reluctant to adopt the cloud platform. For instance, researchers in (Kuo, 2011; Schweitzer, 2012) are more sceptical about the use of cloud services given the ratio of successful implementations of cloud in healthcare systems so far. However, improvements are being made and awareness is spreading about the benefits of cloud enabled architectures among providers and consumers. Moreover, economic analysis of cloud services has also been performed (Schweitzer, 2012) to ensure that budgetary savings are not over estimated due to the hidden costs of using cloud services.

In healthcare, the use of big data with predictive analytics has a great amount of potential, especially, when paired with cloud-based platforms. Cloud computing is facilitating significantly the patients, physician and doctors in terms of data sharing and availability, regardless of the location of the patient and clinicians. It offers an incredible opportunity to improve services and operational efficiency at the same time. Cloud computing with big data extensively helps in speeding-up the process of record keeping. Furthermore, its ability to process big data helps extensively to speed-up the meaningful insights and hidden trends (Armbrust et al., 2009; Manekar and Pradeepini, 2015).

In general, data is considered as big when the data size becomes so large that it can only be analyzed by being processed by the application of special algorithms to the domain of the data (Mayer-Schönberger and Cukier, 2013). Beyer and Laney (2012) presented a more comprehensive description of big data. Furthermore, other authors (Marshall et al., 2015; Onukwugha, 2015) explained that when traditional data processing methods are not able to process a chunk of information consisting of multiple interrelated variables, then that data chunk is considered as big data. Further, in respect to healthcare, big data can be described in terms of certain characteristics such as variety, velocity and veracity (Connolly et al., 2013; Courtney, 2013). Moreover, pervasive computing along with mobile technologies on cloud platforms is generating enormous amounts of data each day. Structured, semi structured and unstructured healthcare data is being generated at dramatic speed. According to a study conducted in 2012, the size of healthcare data was estimated to be around 500 petabytes and its growth was estimated to reach 25000 petabytes by the year 2020 (Sun and Reddy, 2013). Moreover, each day new inventions are appearing which able to exchange large volumes of digital information. Such large volumes of healthcare data are difficult to process and handle by traditional database tools, therefore, specialized software and hardware interfaces need to be designed to meet the growing needs of big data. This large amount of data contains patterns that can be extracted to yield useful outcomes in the future, specifically for healthcare.

Various analysis techniques can be used to predict related patterns in the given set of data. By proactively identifying facts about patient healthcare data, timely preventive measures can be taken which could lead to better outcomes and thus better life style for patients. Moreover, data analytics can provide better support for processing patient data when making health related decisions. Furthermore, by utilizing big data analytics, physicians can prescribe adequate treatment options for their patients in a more appropriate way. Moreover, cost analysis for the suggested treatments will give estimations beforehand to to enable the selection of the most cost effective treatment. Studies have presented the limitations of providing accurate predictions on smaller datasets (Brill, 2003; Halevy et al., 2009). Here, the authors presented the efficacy of applying trained statistical models on big data in order to perform predictive analysis with great precision and unmatched accuracy.

Furthermore, there are a number of techniques to view and analyze big data. Here, dynamic simulation modelling (DSM) is frequently used to process collected data for making predictions regarding the planning and implantation of healthcare systems. For this, DSM uses multiple factors regarding patients such as patient preferences, patient outcomes and performance of healthcare systems. Moreover, DSM has its application for designing formal models showing mathematical representations. These models provide an ability to test healthcare interventions and their corresponding results or outcomes over a certain period of time. Moreover, these models can play a vital role in improving healthcare policies and medical practices by designing big data products in a more innovative fashion. Moreover, DSM helps to link important aspects of healthcare delivery provided by big data giving a wealth of evidence for decision making for healthcare management, policy implementation and logistics at all levels of the organizational hierarchy (Marshall et al., 2015).

Extending to the field of big data, researchers consider data as a raw material which can be processed and refined to add value to existing systems (Mayer-Schönberger and Cukier, 2013; Murdoch and Detsky, 2013). Here, it is described that big data quality and accuracy is a significant challenge when collecting data for specific purposes. It cannot be assumed to be of highest quality (Matthews et al., 2014). Here, the focus of study was to analyze authenticity and clarity in the data being collected where authors suggested the use of multivariate analysis to identify subgroups of patient populations who respond to treatment in different ways. Here, a combination of different datasets were used on the basis of genetic information, electronic health records, and administrative medical information. Further to this, the limitations of dataset size were described and suggestions were made to only perform analysis at a population level where sufficient health data is available. However, it did not give any information about the criteria needed to select variables from a dataset in order to reveal important facts about the big data.

Moreover, the work of Global pulse (Letouzé, 2012) is worth mentioning regarding the use of big data mining to improve healthcare facilities in under developed countries. They had an innovation lab to research emerging approaches for performing analysis on digital data. Here, the main objective was to identify and avoid vulnerabilities in the data. Furthermore, researchers assembled and used free open sourced technologies and tool-kits to analyze data. Based on results, they proposed an integrated network of Pulse labs globally to conduct studies at country level. A more detailed work of Global pulse was published in a white paper named as "Big Data for Development: Challenges & Opportunities" (Letouzé, 2012). Here, the authors described future benefits which could be achieved in developing countries by adopting big data technology. Further to this, they studied early warning systems to detect emergency situations and problems related to the use of digital media. In addition, their objective was to design policies and programs for getting a more detailed picture of reality with real time feedback about program failures. Similarly, a number of other systems have been implemented on the basis of data analytics to harness the benefits of modern technologies so life chances in developing countries can be made equal to that of the developed world. For instance, clinical decision support applications were proposed by authors to make important health decisions based on the suggestions of physicians (Garg et al., 2005). Further to this one study, (Anderson, 1997) proposed an alert system to assist technicians in the event of abnormalities in laboratory operation. Similarly, doctors can be reminded to issue important instructions regarding preventive measures to avoid medical problems for patients (Hersh, 2002). Following the same concept, warnings and reminders about compliance with necessary medical guidelines can also be sent to clinicians to avoid malpractice (Chaudhry et al., 2006).

Another piece of work (Sun and Reddy, 2013) describes an associated factor with the introduction of big data analytics in healthcare. Here, authors explained that complexity with handling data has also increased with the growing size of data. Traditional databases and mining approaches are no longer supported. Hence, new tools and technologies are designed to minimize processing and query time on patient data. Such as Google big query and map reduce. Here, Google big query is used to store and query large datasets in a

Google cloud environment within a few seconds. It provides a super-fast query mechanism to bring efficiency in working. Map reduce is a software tool to process data on big clusters with an ability to execute processes in parallel. It comprises of two parts, i.e., map and reduce. The map function splits work on multiple nodes, then, the reduce function collects the processed results to produce a single set of information. Furthermore, evidence based healthcare services are now being introduced to replace trivial medical practices. Data has become more heterogeneous and drawing data correlations to get meaningful insights has become tedious.

It should be noted here, how useful information buried in data could be, when analyzed carefully for a social good. Moreover, smart decisions made on the basis of extracted facts from this data will yield more precision in diagnosis and treatment. Furthermore, ever increasing healthcare costs and medical services have been the main challenge for nations as a whole and by reducing the risk of re-hospitalization, the associated cost will also reduce. For this, the heritage provider network (HPN) conducts studies and competitions to predict and identify ways to avoid unnecessary re-hospitalization on the basis of available patient data. Once identified, healthcare providers can introduce more cost effective plans for patients to avoid emergencies (Sun and Reddy, 2013). To gain these benefits, data needs to be stored in a systematic manner in such databases which are sufficiently scalable to deal with large amounts of medical data. Based on mentioned challenges and benefits, one can see a promising future for big data analytics in healthcare specifically for rural and under developed areas.

3.4.3 Internet of Things (IoT)

Cloud computing in addition to internet-of-things (IoT) can play a vital role in conducting healthcare surveys by reducing the time and cost associated with data collection (Herb et al., 2012). Furthermore, with the reduction in prices of mobile phones and almost universal coverage, it can make healthcare more accessible at low cost by taking geographical distances out of the equation (Herb et al., 2012; Velthoven et al., 2013).

The future of internet of things (IoT) and cloud computing with ubiquitous healthcare systems (Atzori et al., 2010) has been seen as an emerging field. Work in (Pang, 2013) highlights the major application areas of IoT in healthcare such as remote assistance to the elderly, health monitoring, chronic disease management and fitness programmes. IoT can also be used for making efficient use of limited resources by applying scheduling strategies to healthcare. Furthermore, one study (Vasanth and Sbert, 2012) was focused on modern trends in healthcare IT. Here the authors, described how to achieve cost effective interactions by using secure connectivity between healthcare providers. The author further elaborated on wireless technologies and their possible uses, such as management of chronic disease, monitoring, emergencies and diagnosis of diseases in healthcare networks. Further to this, details were given about infrastructure for on demand delivery of health services comprising of servers, gateways, routers, and databases.

Furthermore, IoT provides a ubiquitous platform to utilize body area sensors to collect data and upload this to cloud servers. Sensors situated on any electronic device (e.g. Bluetooth on a smart phone) can be used to measure the physiological parameters. Similarly, a number of healthcare applications are designed for mobile platforms such as Android, iOS and Windows based smartphones. These operating platforms provide the support for sensors to be used as data collection interfaces in daily life. Healthcare data collected by using these interfaces unveils important facts and patterns about serious medical conditions. For instance, a home monitoring system was developed to take care of elderly people (Luo et al., 2010; Nussbaum, 2006) which ultimately helped to reduce their hospital readmissions. Such systems will also aid in reduction of costs by providing early treatment and interventions. Further studies have shown the application of IoT along with multimedia technologies in the medical field, such as ambient-assisted living (AAL) and telemedicine (Zhang et al., 2010). A more robust overview of smart devices, mobile health, and cloud services in relation to efficient, cost effective, and timely access to health services was given in another study (Gachet et al., 2012; Kuo, 2011). Here, the authors described the modern innovations and systematic approaches being used in healthcare to improve quality of services. Moreover, the generated data from these systems has to be managed properly and analyzed systematically in order to make life less disease prone (Doukas and Maglogiannis, 2012). For this, cloud computing provides comprehensive support regarding control, infrastructure and privacy (Alagöz et al., 2010; Löhr et al., 2010). Furthermore, IoT presents a state of the art solutions for managing medical sensor data (Doukas and Maglogiannis, 2012). Likewise, IoT together with cloud services, makes the handheld device a perfect choice for healthcare information exchange, access and communication (Nkosi and Mekuria, 2010). Further to this, it provides increased data security, availability, and redundancy (Alagöz et al., 2010; Kuo, 2011). Moreover, it has relieved pressure on devices having limited processing power and computation ability by running resource hungry algorithms on cloud. Furthermore, it also saves energy for devices with small power banks (Nkosi and Mekuria, 2010).

Moreover, the concept of smart grids is of worth important mention here. Smart grids provide the functionality for data collection by various systems using IoT technologies and make this data available on-line. Further, the collected data is used in data mining to predict useful patterns in the data for a given domain to control energy consumption and relevant infrastructure cost (Liu et al., 2011). Integration of smart devices with existing systems has the potential to enhance system reliability, efficiency and flexibility on a larger scale with an optimal energy consumption. Moreover, it will enable policy makers to plan for required resources in terms of cost and manpower. Darianian and Michael (2008) and Yang et al. (2014) described the application of IoT in various domains to improve daily life. Here, they suggested the use of sensors and actuators in home and office buildings for surveillance and monitoring purposes. Further to this, extending the studies in this domain, other authors (Schaffers et al., 2011; Vicini et al., 2012) suggest how to create smart cities by using. Moreover, IoT technologies at a broader level help control traffic, monitor parking spaces, even providing trash notifications when containers are full in a specific area. Here, the suggestions of authors seem worthwhile but application brings associated challenges with the possible need for redesign of the architecture of cities.

Moreover, dealing with personal privacy, is another challenge. As traceability of objects increases, personal data regarding daily activities will become more vulnerable to privacy threats. Further to this, it also has related social consequences as well. For instance, as with the society of Pakistan, which is uncomfortable about sharing personal data on a public platform and has a complete set of acceptable norms to control this kind of sharing. Following on this same discussion, assurance of personal privacy is a core value that is needed for society to progress further down the road in the implementation of IoT. Here, one way to ensure privacy is to implement privacy and data ownership policies. In this way, smart devices will require authentication to use and permissions for other devices ton interact with them (Roman et al., 2011).

Furthermore, to avoid the dominance of a single group in power, a universal legal set of policies are needed. To gain full benefits from IoT smart objects, a stakeholder based shared governance structure is required on a broader level (Weber, 2011). Moreover, in case of any violation of legal constraints, a global accountability and law enforcement check is mandatory for effective governance (Weber, 2011).

Moreover, work by some authors (Alemdar and Ersoy, 2010; Ko et al., 2010) related to wireless sensors network (WSNs) can be seen as pioneering for healthcare research activities in this field. Contrarily, the trend has been shifted to IP-based sensor networks called IPv6-based low-power wireless personal area networks (6LoWPAN). To further extend in the domain of WSNs in relation to IoT, a more detailed explanation is given by a group of different authors (Alcaraz et al., 2010; Christin et al., 2009; Mainetti et al., 2011). More details about WSN based healthcare services are described by Chung et al. (2008), where the application of WSNs to healthcare services with IoT is studied. Contrarily, wearable devices provide heterogeneous features with IoT in the healthcare domain. However, this heterogeneous nature of wearable products and medical sensors raises different challenges in their mutual integration. In another study (Castillejo et al., 2013), details of applications based on the integration of WSNs and wearable devices together with the corresponding outcomes are given. The said combination opens up an innovative field with a wide variety of mobile based health applications in addition to other smart devices such as smart watches.

For instance, an activity recognition system for mobile devices to monitor patient activities remotely was presented in another study (Sebestyen et al., 2014). Another similar system based on IoT for remote monitoring with wearable devices was explained by researchers (Bazzani et al., 2012). Moreover, Bluetooth low energy (BLE) and its combination with WBAN for wearable devices was studied in detail in another paper (Jara et al., 2013).

Here, it is worth mentioning the concept of IoT healthcare network (IoThNet) which attains a core importance in healthcare IT. It provides access to the IoT backbone and helps in the exchange of medical data in both a convenient and secure way. Furthermore, authors in another study (Gronbaek, 2008; Zhu et al., 2010) provided a detailed overview of proposed architectures in IoThNet. Similarly, further studies (Imadali et al., 2012; Istepanian et al., 2011; Yang et al., 2012; Zhao et al., 2011) entailed various conceptual deployment structures related to IoThNet. Here, IoThNet gives insights about computing platform as well as network platform employed in designing healthcare systems. More details are presented (Wang et al., 2011), proposing a framework for the health information of residents using a service platform. Here, the authors highlighted the ability of caregivers to communicate with application databases through a support layer. In the same manner, Yang et al. (2014) suggested the use of data center platforms as middleware between the business logic layer and smart objects. The concept of developing open platforms for all stakeholders by following a standardized approach for designing interfaces, both software and hardware, was presented in (Pang et al., 2014). Author further suggests the possible interface types or categories, electronic health data, and security configurations. Here, the main concept behind open platforms is to enhance collaboration and interoperability. Another, related approach in IoThNet platforms is automating design methodology (ADM) platform. The main applications of ADM is for rehabilitation purposes (Fan et al., 2014). The details about event driven middleware package named as VIRTUS was given by Bazzani et al. (2012), describing the use of XMPP instant messaging protocol for secure communication in IoThNet. This approach could be appropriate in remote areas where network connectivity is weak or fluctuating.

Rasid et al. (2014) presented an approach to handle communication between multiple users and multiple sensors. Here, a gateway is used to read health data from routers on the edge. After reading, the captured data is parsed by applying certain algorithms according to a defined format. Xu et al. (2014), proposed a cloud enabled system to access healthcare data over IoThNet. Further to this, the authors describe the 3 layer architecture consisting of the database layer, resource layer and business logic layer. Here, the database layer supported the storing of data in multi-tenant databases whereas the resource layer managed the data access mechanism and business logic layer and was responsible for data sharing and other operations. Further to this, the important fact to be mentioned here is that the distributed healthcare data was managed and organized by a resource controlling approach.

Moreover, despite the added benefits of smart objects and IoThNet in improving healthcare facilities, none of the system is able to provide specialized and satisfactory services to incapacitated and ageing individuals. Shahamabadi et al. (2013), presented an artificial intelligence based IoT environment called ambient assisted living (AAL) to ensure quality healthcare delivery for elderly people. AAL was proposed to deliver to the elderly the convenience of living an independent life with less monitoring and in a secure environment. AAL serves the elderly with a human like service quality able to deal with any unwanted situation, thus restoring their confidence in living according to their own liking. Further to this, the author gave major insight into AAL regarding the architecture, security control and communication mechanisms for the proposed IoT model. Similarly, G.Zhang et al. (2012) studied cloud based platforms. Here, the authors strived to develop closed loop healthcare services with smart object KIT protocols. Furthermore, authors implemented a gateway on a desktop machine to verify different aspects of the proposed platform. The observed aspects included quality of service (QoS), interoperability, privacy and security, data storage and the feasibility of implementing in a given environment. Further studies highlighted the required technological support for AAL with a proposed architecture to implement these technologies (Istepanian, 2011). In addition, AAL based medication control is presented by other researchers to analyze secure services for IoT (Goncalves et al., 2013).

Prasad and Prasad (2011) presented a cooperative platform for a rural healthcare monitoring system. Here, monitoring services were provided by utilizing the IoT concept within a local community. A network was established connecting a local hospital and nearby rural residential surroundings and a specialized healthcare service was provided called community healthcare. Authors analyzed this system and found it to be energy efficient. Further enhancements can be made by applying authentication of users over the cooperative network. You et al. (2011) employed similar concepts where researchers proposed a community medical network consisting of wireless body area networks (WBANs). This network can be regarded as a virtual hospital or clinic. Furthermore, a four layer information service platform is proposed for exchanging medical data between healthcare entities and service providers. This system was proposed for specialists to gain useful medical suggestions by specialists remotely (Wang et al., 2011).

Similarly, sharing huge amounts of health data by applying certain semantics and ontologies was described by some researchers (Burgun et al., 1999). Medical ontologies and semantics in healthcare applications based on IoT have become a core value in health IT. On the basis of these semantics and ontologies, a separate service has been introduced named as semantic medical access (SMA). Work in G.Zhang et al. (2012) present a semantic medical monitoring system based on IoT sensors. Here, sensor data is saved by cloud and healthcare applications using medical rule engines to analyze data in the cloud. Other researches (Miori and Russo, 2012; Xu et al., 2014), described various approaches employed to collect and integrate IoT data related to emergency health services. Further to this, authors (Jara et al., 2013; Jia et al., 2012; Sebestyen et al., 2014) entailed possible challenges and issues related to IoT platform. Another group of researchers (Liu and Yang, 2011; Xiao et al., 2013) provided a detailed overview of emergency situations with appropriate scalable solutions to overcome the corresponding challenges. Here, emergency situations may include accidents, earthquakes, fire and weather disasters. For these, a dedicated indirect emergency healthcare service can be provided with an ability to send notifications, alerts and emergency suggestions. The author further argued that the said issues have been neglected and were unable to attain the attention of researchers on a larger scale.

Istepanian (2011) presented a real time glucose sensing system for the condition of diabetes which was based on IoT. The proposed system monitors the changes of glucose levels in the blood and identified change patterns which were ultimately found to be useful for managing diet and medication plans. Further, the authors described the connectivity of these sensors to healthcare providers. Similarly, another blood glucose somatic data collection model was proposed by Guan (2013). Here, an IoT based transmission device was used for transmitting data on network. The model included a blood glucose collector device, a mobile phone/computer and a processor. Furthermore, a system for medical acquisition detection was described by Lijun (2013). The authors considered it to be more of a generic approach based on IoT to monitor glucose levels in patients. A relevant work is given in a paper by Wei et al. (2012). Moreover, diagnosis of various heart related anomalies and diseases by monitoring the ECG of patients was performed in the work presented by Drew et al. (2004). Here, they monitored rhythmic patterns in the heartbeats of patients. In addition, application of IoT based systems to examine heart related anomalies was detailed in various other studies (Agu et al., 2013; Castillejo et al., 2013; Dash et al., 2002; Ge et al., 2014; Jara et al., 2013; Rasid et al., 2014; Yang et al., 2014; You et al., 2011).

Furthermore, Liu et al. (2012) evaluated an ECG monitoring system which had a portable acquisition transmitter and a receiving processor with an ability to communicate wirelessly. Here, to detect abnormality in heart data, a real time search automation approach was used. Moreover, an ECG signal detection algorithm was also evaluated and examined by Y. Xiaogang and Wentao (2011). This algorithm was designed to be employed at application layer within an IoT network model. Similarly, to monitor blood pressure of patients, a monitoring system based on IoT was proposed by Dohr et al. (2010). The presented system consisted of a KIT blood pressure meter and KIT mobile phone having an NFC interface installed. Moving in the same context, Sebestyen et al. (2014) presented a remote blood pressure controlling scenario where they suggested a communication model for healthcare units. Further to this, Tarouco et al. (2012) described a blood pressure device designed to work with an Apple mobile phone on an iOS platform. Similarly, to exchange blood pressure data over an IoT network, Guan (2013) proposed a mechanism for a specialized device comprising of an apparatus for blood pressure monitoring and a communication module. Moreover, Qi et al. (2012) studied a blood pressure monitoring terminal with location tracing functionality.

Another application of IoT in healthcare is to maintain homeostasis (i.e. to keep body temperature at optimal level) which was explained by Ruiz et al. (2009). Furthermore, using temperature sensors to monitor body temperature within a mobile IoT (m-IoT) system was presented by Istepanian (2011). Here, various samples of body temperatures were used to verify the presented approach. Related to this, Jian et al. (2012) examined a home gateway with IoT for temperature controlling purposes where the authors described the use of infrared radiation as the communication channel to transmit the temperature values of users. Furthermore, an RFID based temperature sensing and controlling system was studied by In (2014).

Moreover, to monitor blood oxygen saturation (i.e. Pulse oximetry) levels, IoT based systems are of much importance in healthcare. Work by Khattak et al. (2014) entailed core details of IoT applications in pulse oximetry. Here, a survey of healthcare services with an integration of IoT was presented comprehensively. Extending further, Jara et al. (2013) described an innovative wearable pulse oximeter with Bluetooth connectivity feature. Here, a sensor used was connected to the Monere platform directly. Larson et al. (2012) studied the challenges of cost and power consumption regarding sensors to monitor pulse oximetry. The authors then suggested a state of the art cost effective and low power oximeter to provide assistance to remote located patients. The proposed oximeter was able to provide continuous assistance to users over an IoT network. Larson et al. (2012) described telemedicine applications along with proposed pulse oximeter in detail. A wearable oximeter based on WSN is explained by Larson et al. (2011). Here, authors argued about the possibility of adapting WSN for an IoT network for the desired functionality.

Moreover, IoT has a true potential to overcome the challenges of healthcare delivery in rehabilitation systems. In addition, it has produced a large set of benefits in areas where there is a shortage of health care specialists. Fan et al. (2014) recommended an intelligent rehabilitation system using an ontology based design. They further put forth the impact and related positive outcomes of IoT to connect all of the required resources for real time information reporting. This system can be used to help elderly people facing aging problems and challenges. The work presented by Tan and Tian (2014), further elaborated the fact of remote rehabilitation services by implementing IoT technologies. A different case was studied by (Lin, 2013) to provide rehabilitation healthcare facilities inside prisons for prisoners. Further, the authors explained the rehabilitation system and its integrated architecture. Work by Liang et al. (2011) described the childhood autism condition and the proposed training to overcome language barriers. Another study (Y. Yue-Hong and Yi, 2014) presented the emerging concept of smart city and collective health rehabilitation system.

To avoid discrepancies in medication management, an intelligent packaging approach for medicine boxes was presented by Pang et al. (2014). Here, two prototypes were studied namely, iMedBox and I2Pack. Field trials were carried out and packaging activity was monitored and communicated wirelessly. The objective of the presented system was to ensure the intact sealing of medicine boxes to restrict possible threats to public health. A related system was also examined by other authors (Laranjo et al., 2013), where RFID tags were used for controlling medication. Here, the communication channel used was provided by an IoT network. Moreover, the objective here was to design a medication control system to aid AAL solutions.

A unique work was done by researchers (Yang et al., 2014) where they proposed smart wheelchair management system based on IoT. The proposed work has enormous advantages for physically impaired persons. The system comprised of sensors and WBANs and communication was done over an IoT network. Another study (Kolici et al., 2014) proposed a system with a peer-to-peer communication mode for a medical support system. The system had functionality to control wheelchairs by a vibration control mechanism. A comprehensive study was done to manage wheelchairs by Intel (Dr. Hawking's Connected Wheelchair

Project). Here, research entailed that connected device can be controlled by data. These devices had the ability to monitor vital signs of the patients using wheelchairs by collecting data from smart surroundings.

In addition to these facts, there are number of other healthcare applications that utilize wise medical devices with IoT to aid diagnoses, laboratory testing and decision making for patients. The most prominent and widely studied areas for designing such applications are surgery, detection of hemoglobin, abnormalities in expiratory flow, abnormal cellular growth, treatment of cancer, eye disorders, and in dermatology to name a few (Gayat et al., 2011; Pesta et al., 2013; Prasad and Prasad, 2011). The authors further explained various types of diagnostic tools and devices together with their methods of communication and connectivity. Generally, these healthcare devices are controlled by mobile applications to provide services on the go, anywhere and anytime. To harness the full power of sensor equipped handheld devices along with IoT, state of the art hardware and software systems have been implemented. Mosa et al. (2012) provided a comprehensive survey of healthcare applications that are adding value to life. Here, the presented survey included an extensive overview and detailed discussion regarding patient needs, health education, required training, and other health information or awareness applications. Furthermore, there are numerous such healthcare apps available that are developed to meet user needs according to specific nature of the requirements.

In general, studies have effectively analyzed the pros and cons of cloud computing together with IoT. Here, various aspects of these technologies were presented such as storage, interoperability, security, communication and quality of healthcare service with related shortcomings and much needed improvements to make the most out of these, specifically for underdeveloped regions like Pakistan. Moreover, in healthcare, cloud computing is facilitating many health applications, including deployment of EHR, data sharing and enhancement, management of big data, patient enrollment, revenue cycle management, and claims processing, just to name a few. With the passage of time, cloud computing is gradually becoming more widely adopted by healthcare organizations across the globe. It requires major consideration in terms of understanding possible benefits and the risk factors associated with achieving them. Different models of services such as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) and deployment models such as private, community, public and hybrid also need attention to provide adequate scalability (Priyanga and MuthuKumar, 2015). Though, providing new and improved patient care capabilities, challenges that accompany the implementation of cloud oriented systems architecture include data privacy and security, data portability, and reliability.

3.5 Intelligent Computation Techniques

Various Data mining and knowledge discovery techniques can be used for modelling and classifying population based data in order to identify patterns related to different levels of health-shock suffered by individuals as a consequence of socio-economical, environmental and health related factors. Techniques such as clustering, classification, regression, association rule mining and decision trees have been widely used in the health-care domain (Shukla et al., 2014). In one study (Choi et al., 2010), six classification algorithms namely decision trees, artificial neural networks, logistic regression, Bayesian networks, Nave Bayes and the stacking-bagging method. These algorithms have been used for building classification models for helping public health-care centres to plan and implement health-care service programs based on the health-care demands of local residents. A study was presented in a paper for predicting cardiovascular autonomic (CA) dysfunction in the general Chinese population using artificial neural network (ANN) based prediction models (Liu et al., 2013b). Authors (Yu et al., 2013) have also applied support vector machines (SVM) for classifying and detecting people with diabetes and pre-diabetes in a cross-sectional representative sample of the U.S. population.

Rule induction approaches such as associative rule mining, rule based classification, and rule based fuzzy logic systems have also been successfully applied within the health informatics domain. A data-adaptive rule-based classification system for Alzheimer's disease classification has been designed (Jain et al., 2013). Here, the system generates relevant rules by finding adaptive partitions using gradient-based partitioning of the data. The adaptive partitions are generated from a histogram that is used for analyzing the discovered rules, which are used to assist in classifying the new data correctly.

In another paper (Fleming et al., 2007), a fuzzy based approach was presented for cholera prediction based on a case study of Southern African. The proposed approach takes into account various factors including preconditions for cholera outbreaks, environmental conditions, and socio-economic factors for building the prediction model. Here, the fuzzy model was developed based on historical information, expert knowledge, and climatic and biophysical parameters while cholera outbreak risk was the output parameter. The approach is aimed at helping to minimize the impact of cholera by informing the policy makers using a fuzzy prediction model. Similarly, other applications of fuzzy systems have been used to manage malaria (Djam et al., 2011), conduct medical diagnostics (Innocent and John, 2004) model epidemiology to risk factors (Massad et al., 2008) in addition to cancer treatment (Poongodi et al., 2012) and the prediction of water shortages(Altunkaynak et al., 2005). In another paper by (Doctor et al., 2014), an adaptive fuzzy rule based linguistic summarization (LS) system is proposed for modelling the behavioural cues of dementia patients based on monitoring their interactions in the home through the use of smart devices and environmental sensors. Here, the aim is to use the generated rules (weighted according to the profile patterns of the behaviour of an individual) to track behaviour changes that could indicate cognitive decline owing to disease progression. Generally, fuzzy logic systems have proved to be an ideal choice to model health-care systems due to their ability to handle uncertainties, imprecisions, complexity and incompleteness of information (Szolovits, 1995; Szolovits et al., 1988). Moreover, fuzzy systems provide transparent and flexible rule based models through the use of linguistic quantifiers (Mendel, 2001). These models can provide a methodology for predictive modeling and classification using approximate reasoning of uncertain information.

Various studies regarding LOS prediction and healthcare models will enable early forecasting to prevent future complications and increased LOS. Moreover, efficient utilization of resources and healthcare facilities using these prediction models could be important for decision making. To find out the probability of re-hospitalization for patients at high risk, the case of patient at risk of re-hospitalization (PARR) was studied by researchers (Billings et al., 2006). Here, prediction algorithms were designed to predict the risk score of re-admission on the basis of previous discharge data of patients within the one year of their initial disease identification in real time. In another paper (Houthooft et al., 2015), researchers put forth various machine learning approaches and techniques to predict the patient length of stay (LOS). Here, these techniques were trained to calculate scores for sequential organ failure (SOFA) on the dataset consisting of 14,480 patients. Furthermore, with regard to patient LOS and forecasting the patient bed requirements within a healthcare unit, the Henry Ford Health System (HFHS) was developed (Meitzner and Engineer). It could prove beneficial to meet and predict the future needs of budgeting, staffing decisions and procurement strategies for increased healthcare efficiency. Moreover, an ensemble of classifiers-SVM, ANN and decision trees were iterated (Stoean et al., 2015) to develop a more accurate prediction model using a voting scheme and accuracy in discriminative guidelines through visual rules. This study was conducted to estimate the LOS of post-surgery patients with colorectal cancer. Furthermore, Multiple standard classification algorithms were used together (Sheikh-Nia, 2012) to predict LOS effectively using multi-dimensional medical data. Here, two sequential ensemble classification (SEC) schemes were proposed, the first based on an ensemble of homogeneous classifiers and the second based on a heterogeneous ensemble of classifiers, in three hierarchical levels. A combination of employed algorithms is to increase performance for complex patient data. To explore factors associated with suspected community-acquired pneumonia (CAP) requiring hospital re-admissions, different classifiers and regression models (Huang et al., 2006). The identified symptoms—shaking, sweats, weight loss, functional impairment, renal and heart diseases were determinants of LOS prediction.

Another study regarding LOS predication was carried out at Boston Medical Center by the trauma surgeons. Outcomes were calculated for drivers of more than 70 years of age or younger, using multiple linear regression (LaMorte, 2015). Another methodology (Azari et al., 2012) suggested clustering for training set creating and to train classification algorithms for predicting hospital length of stay (PHLOS). Here, performance of each training set was analyzed for LOS predictions. Similarly, other work (Combes et al., 2014) uses predictive modelling techniques in a proposed decision support system (DSS), based on patient data to support increasing demand for emergency medical care. Further in case of Hemodialysis, another study (Yeh et al., 2011) presented the fact that hospital re-admission rate of hemodialysis center was reciprocal to the healthcare service quality. Here, data mining techniques for biochemical data of the subject patients was analyzed to develop a decision support system and the extracted temporal patterns from this system were considered to be helpful to predict hemodialysis re-admissions.

Makar et al. (2015) tried a wide array of machine learning classifiers for mortality prediction of the elderly by using an administrative claims database. The study mainly focussed on improving decision making for high risk patients later in life. In another study (Walsh et al., 2004), a neural networks technique was used for pattern recognition for Bronchiolitis disease in infants. For this, analysis was performed on specifically designed database to test, validate and train a neural network ensemble.

Similarly, to predict post-hospitalization venothromboembolism (VTE) disease, researchers (Edelsberg et al., 2006; Kawaler et al., 2012) applied classification models to predict the possibility of contracting VTE. These models were made by considering health variables such as vital signs, demographics, medications, genetic markers and other personal health records maintained using EHRs. Here, machine learning techniques were used on these models to aid post hospital patient management.

3.6 Why Model Healthcare Data with Fuzzy Logic?

In the era of internet-of-things and big data, healthcare poses unprecedented opportunities for cloud computing, machine learning, data mining, and evolutionary computing (Massad et al., 2008). However, there are inherent complexities, uncertainties, and imprecisions involve in healthcare delivery and recording (Abraham and Nath, 2007). For instance, diagnosis and management of disease is complex due to the large number of variables involved.

Moreover, due to the low literacy rate in developing countries, patients cannot describe exactly how they feel and on the other hand, doctors and nurses cannot exactly characterize what they observe (Szolovits, 1995; Szolovits et al., 1988). Similarly, carelessness of technicians and/or malfunctioning of instruments, makes the laboratories results error prone. All the complexities of medical practice make traditional quantitative approaches of analysis inappropriate (Szolovits et al., 1988).

Here, it is worth mentioning that domains like healthcare require a lot of expert opinion and subject matter advice. Humans, naturally solve problems with a lot of factual information and the ability to reason with that information (Massad et al., 2008) makes expert systems, the ideal choice. Expert systems are based on two major components, a knowledge-base (i.e. collection of specific data and rules) and an inference-engine (i.e. having reasoning ability to come up with conclusions based on the knowledge base).

Moreover, among all evolutionary computing methods like neural networks, genetic algorithms, particle swarm optimization, gradient decent and many others, fuzzy logic proved to be an ideal choice to model healthcare systems due to its adaptability with uncertainties, imprecisions, complexity, and incompleteness of information (Massad et al., 2008; Szolovits, 1995; Szolovits et al., 1988). Moreover, it resembles human decision making, i.e., ability to work from approximate reasoning and ultimately finding a precise solution.

3.7 Summary

This chapter has discussed the health scenario of underdeveloped and developing countries in addition to the importance of the dataset that is required to understand health-shock and the causes. Furthermore, it discusses a number of intelligent computational approaches in the light of inherent complexities, uncertainties, and imprecisions associated with healthcare delivery and recording. Moreover, it is highlighted that among all evolutionary computing methods like neural networks, genetic algorithms, particle swarm optimization, gradient decent and many others, fuzzy logic proves to be an ideal choice to model healthcare systems due to its adaptability with uncertainties, imprecisions, complexity, and incompleteness of information. Moreover, a summary of the literature review is given in Appendix A.1 - Appendix A.3.

As a result of this chapter, it is argued that if the health dataset is collected appropriately using a mobile based cloud infrastructure and analyzed using fuzzy logic based approaches, it will provide insight into the health dataset needed for predicting health-shock. Furthermore, such analysis will help to reform healthcare policies in order to mitigate health-shock and the causes in addition to initiate community based health programs.

Furthermore, a discussion about the invalidity and uncertainty in survey based data collection is explicitly performed in this study. There are a number of data modelling techniques which are highlighted in the literature. Here, a fuzzy approach for uncertainty in the dataset collected is viewed in terms of privacy issues which are tackled by applying data encryption techniques. Here also, fuzzy systems employed linguistic quantifiers to provide transparent and flexible rule based models for predictive modelling and classification using approximate reasoning for uncertain information. The literature further highlights that fuzzy logic could be an effective technique to handle uncertainty in the data and to model the survey data for health shock prediction due to its flexibility and more human like decision making.

It should be noted how useful this data could be when analyzed carefully using these prediction models to provide information for important decision making. However, one of the major limitations is the lack of an available comprehensive healthcare dataset that could provide prediction of health-shock, especially in the case of Pakistan. In this regard, the research presented here is a pioneering contribution to fulfil the requirement for Pakistan's first healthcare dataset. Moreover, predictive analysis can be performed on this survey based dataset to provide an insight into people's health.

In conclusion, this chapter has effectively presented and analyzed the benefits and limitations of various systems used in the healthcare domain. Here, cloud computing in addition to IoT and big data technologies were studied giving major consideration to factors such as storage, interoperability, security, communication and quality of the healthcare service with discussion of the related shortcomings and much needed improvements, specifically for underdeveloped regions like Pakistan.

The next chapter will present the research methodology in addition to the details of Pakistan's first context-aware healthcare dataset in order to understand and monitor health-shock in Pakistan.

Chapter 4

Data Collection

4.1 Introduction

To understand health-shock and its causes in the rural and remote areas of Pakistan, a dataset of 1000 households from the district Haripur has been collected. This dataset contains 47 features including age, job status, education, gender, living standards, frequency of major and minor illness, sewerage system, and access to basic amenities of life, just to name a few. One of the main objectives of this particular survey was to identify the shortcomings of the present healthcare system in Pakistan—especially, in rural and tribal areas. Furthermore, it highlights the difficulties faced by people living in the rural areas of Pakistan.

The schema of this chapter is as follows: Section 4.2 discusses the data collection activity; Section 4.3 provides an overview of the District Haripur (research settings); Section 4.4 explicates the collection, cleaning and staging of the dataset, along with discussing the ethical considerations, reliability and validity, and survey questionnaire; Section 4.5 discusses the data coding and quantitative analysis, quality assurance measures and pilot study; and Section 4.6 provides a summary.

4.2 Data Collection

To understand health-shock and its causes in the rural and remote areas of Pakistan, we collected a dataset of 1000 households from the district Haripur with the help of Begum Memhooda Welfare Trust (BMWT) which is a non profit organization working on five major programs, namely: health & medical support, education support, food support, wedding composition support and income support in Haripur district since June 2008 (Details are given in Appendix D).

Owing to the trust of the local community in the BMWT, we were able to collect features like age, job status, education, gender, living standards, frequency of major and minor illness, details of the sewerage system, and access to basic amenities of life, during the survey. Here, it is worth mentioning that the survey questionnaire was designed by gaining insight from various available datasets (GoP, 2012, 2015; Naveed and Islam, 2012; Schreiner, 2010). Moreover, as per the consent form (see Appendix D.1), all of the personal information of the responders was kept confidential, i.e., responders' identity information and other unique features that could help to identify the responders were kept anonymous by BMWT before sharing the data for this research.

Here, one of the main objectives of the survey was to identify the shortcomings of the present healthcare system, especially in rural and tribal areas of Pakistan. During the survey, data from 1000 households was collected from 29 villages in the proximity of district Haripur by using the survey instrument as shown in Appendix D.2. Here, the survey instrument was designed by following the rules mentioned in (Nations, 2005). Fig. 4.1 presents the geo-coordinates of the villages, which took part in the survey activity.

In order to provide a clear picture of Haripur district, an overview of this rural area is in Section 4.3.

Figure 4.1: Geo-coordinates of the villages.

4.3 An Overview: District Haripur, Pakistan

Haripur district is in the Hazara region of Khyber Pakhtunkhwa province of Pakistan. It is located in a hilly plain area at an altitude of around 610 metres (2,000 ft) above sea level. According to (GoP, 2010), it has an estimated population of $1,024,497$ with a rural to urban ratio of 88%:12% as shown in Table 4.1.

Table 4.1: Details of Haripur District

Total Households Approx:	146,357
Average HH Size Approx:	7 persons
Male & Female Ratio %	49.92%:50.08%
Population Density	2.22 Persons/Acre 550.26 Persons/sq km
Rural & Urban Ratio	88% : 12%
Literacy Rate Approx:	65%

Appendix B shows the thematic map that represents the population of different union councils of district Haripur. In district Haripur, the monthly income of the family depends on the number of persons who are involved in skilled labour, unskilled labour and/or child labour. Here, it is worth mentioning that only 13.33% households have monthly income of PKR 15000, which is approximately 150 USD. Furthermore, only 3% of the households have a monthly income of PKR 35000, which is approximately 350 USD as shown in Fig. 4.2.

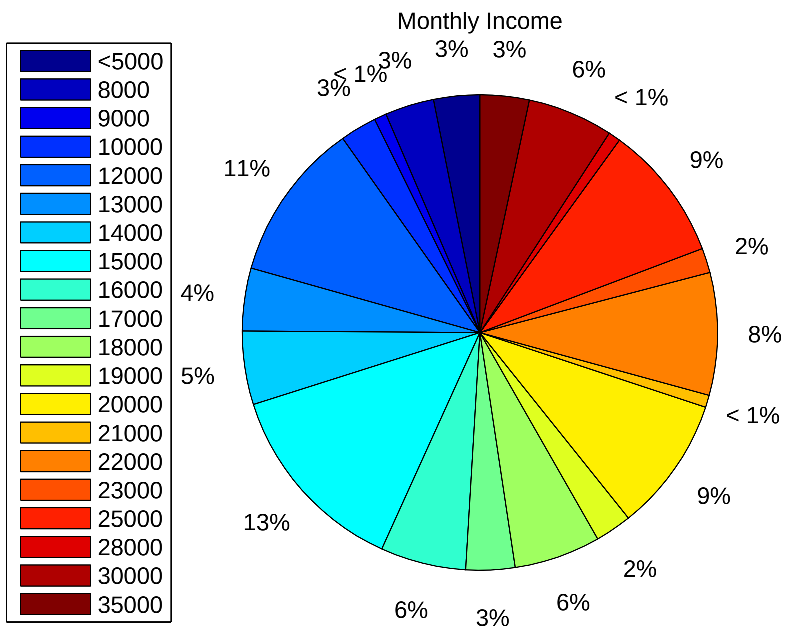


Figure 4.2: Monthly Income of Households in District Haripur.

In district Haripur, more than 42% of the population are below the poverty line; whereas, 31% of the population are in the middle-income group. A few of the main reasons for poverty include: an uneven distribution of resources, poor human resource development and polarization of power. Owing to poverty, one of the main sources of income is child labour as shown in Fig. 4.3.

Moreover, for an estimated population of $1,024,497$, there are about 6 hospitals, 6 rural health centers (RHCs), 6 sub health centers (SHCs), 42 basic health units (BHUs), 2 mother and child health centers (MCHs), 9 dispensaries, 1 tuberculosis (TB) clinic and leprosy clinic. Appendix C shows the plot of all the health facilities on the map of district Haripur.

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

Figure 4.3: Income Sources of Households in District Haripur (GoP, 2013).

Here, it is worth mentioning that due to a shortage of medical doctors and paramedical staff in addition to different medical facilities, the number of patients per month at different BHUs is not more than 300. However, in comparison to BHUs, the patient load at DHQ hospital Haripur is around 15,000 to 18,000 due to the presence of a sufficient number of doctors, paramedical staff and different medical facilities (GoP, 2010). Furthermore, in BHUs, a single doctor is being shared by multiple BHUs. For instance, one single doctor is being appointed for both BHU Amgah and BHU Bandi Sher Khan. Owing to such situations, people of district Haripur have to spend out-of-pocket either to reach hospitals in urban areas or to avail themselves of private health facilities, which act as one of the major reasons for health-shock.

It is obvious from Table 4.2 that the current health infrastructure for the district does not correspond well with its population. In the district of Haripur, there is only one bed for every $2,247$ people. In contrast to Haripur, there is one bed for every 100 people in the

Table 4.2: Bed Count in District Haripur

Sr. No.	Health Facilities	Total no. of beds
1	Hospitals	68
2	BHUs	0
3	RHCs	17
4	SHCs	0
5	Dispensaries	0
6	MCHs	20
7	TB clinic	20
8	Leprosy Clinic	10

developed countries of the world. Moreover, in district of Haripur, IMR is 66 whereas overall MMR in the province of Khyber Pakhtunkhwa is 275 (GoP, 2010).

4.4 Dataset: Collection, Cleaning, and Staging

Estimating the health-shock of a region is a complex phenomenon that involves multiple indicators and associated complexities such as living standards, health risks, access, income allocation, and poverty state of a region. Multidimensional nature of health-shock demands a multidimensional measurement approach to interpret its effects on the living standards of a society. Furthermore, to estimate health-shock, an in-depth study and detailed survey of the region is needed to highlight and understand the complex factors that add to the state of health-shock.

In this regard, to find the ground realities of healthcare facilities in rural areas, 47 features related to living standards, health risks, access, and income allocation have been collected. Some of these features are the age, marital status, sex of the household head, involvement in the labour force, education of children, financial and water resources, access to health facilities, schools and clean water, effects of climatic changes (frequency of minor illness due to changes in weather, resistance of house against strong winds, severe rain, snow or hail), effects of shortage of basic facilities like fuel, food, money for treatment of illness or fertilizer for crops and waste disposal trends.

4.4.1 Ethical Considerations

Based on the cultural norms of the respective region such as caste, language, opposite gender interaction, approval from village elder, fear and risk of being exposed to outside world, we have followed eight ethical concerns during this survey, which includes: risk assessment undertaken by the BMWT for each study, informed and written consent of any and all participants prior to collection of data, voluntary participation, disclosure of patients' rights to refuse participation without any repercussions, disclosure about the purpose of data collection, anonymity and confidentiality of the collected data, non-traceability of the confidential data and review of any compensations to the participants. The participation of the sample population was voluntary, i.e., all the responders participated in the survey of their own free will. Furthermore, to keep the confidentiality intact, names and details of the participants were not disclosed to anyone as shown in Appendix D and Appendix D.1.

4.4.2 Reliability and Validity

Generally, reliability can be described as the degree of consistency of the survey instrument's measurement for which it is designed to measure (AERA, APA and NCME, 2014; Cronbach, 1971). Here, data collection biases were minimized by the researcher in that only one researcher administered the questionnaires and standardizing conditions. Furthermore, to ensure the data validity, the following principles have been followed throughout, that is, a broad sample, emphasize on important features, and appropriate questions to estimate the skills (AERA, APA and NCME, 2014; Cronbach, 1971). The survey questionnaire has also been fashioned in accordance with the respective principles.

4.4.3 Survey Questionnaire

For the purpose of conducting this particular study, a multidimensional survey questionnaire was designed by taking insight from a paper by (Naveed and Islam, 2012). A printed questionnaire was used to obtain the survey information from the respondents directly. Here, a questionnaire was preferred over interviews as questionnaires were more systematic and offered more comfort to the respondents as they can fill it in their own private settings with discretion (Gill et al., 2008).

In this study, questionnaires were also written in the national language of Pakistan, i.e., Urdu. Also, further assistance was provided to the participants, who could not read and/or write. The questionnaire was designed with the aim of obtaining geographic, demographic and socio-economic data, so that a comprehensive picture regarding the living standards of the participants and effects of health-shock could be drawn. Furthermore, as the targeted population was too large to survey, probability sampling was used in identifying the participants. Appendix D.2 shows the multidimensional survey form used for the data collection purpose.

4.5 Data Coding and Quantitative Analysis

Here, for the open-ended questions, a quantitative analysis was preferred which was a formal, systematic, and objective process used to describe and test the relationship between, and the causal interactive effects among variables (Neuendorf, 2002).

Moreover, for content analysis, data coding was preferred involving coding and categorizing the primary patterns in the data (Han and Kamber, 2006; Hand et al., 2001; Hastie et al., 2009). Furthermore, in order to frame the research, a systematic design was prepared which focused on multiple sources of data, appropriate data analysis activities, and a preliminary conceptual model. Here, it is worth mentioning that throughout the study, all the activities including data collection and analysis activities, source of data and the collection method, data preparation/management and analysis, were properly documented.

Thereby, with regard to the above stated discourse of qualitative and quantitative approaches, the quantitative method is considered to be the best approach due to the very nature and context of this particular study. The quantitative method is suitable, because the research affects the populace and is concerned with improving the health scenario of people in the rural areas of Pakistan.

4.5.1 Quality Assurance Measures

For the quality assurance of the questionnaire, different quality measures were adopted from the perspective of the respondent, i.e., the understandability, comprehensiveness, and acceptability of the survey questionnaire. The details of various quality assurance tools (Biemer et al., 2004; Presser et al., 2004), which were used in the survey, are given below:

- Sample Testing − by asking few expert interviewers and respondents to fill out the survey form and incorporating their suggestions.
- Cognitive Testing − by cognitive interviewing of respondents, where they visualize the questions and re-iterate in their own words.
- Behavioural Testing − by making slight changes to the questions and measuring the difference in the answers of the respondent to determine whether the wording in the questionnaire itself is affecting the overall answers?
- Special Probing − by helping respondents by explaining the intent of the questions in the local language, to make the time taken to answer the questions manageable.
- Experts Opinion − by seeking the advice of experts in the field in order to improve the questionnaire content.
- Compare and Contrast − examining the questions and responses from previous surveys to establish best practice, avoid errors and identify any improvements that could be made to the questionnaire.

Here, it is worth mentioning that the questionnaire was also discussed with various stakeholders from government agencies, experts, and NGOs which helped us to form evaluation criteria. Based on the discussions, the following factors were also considered during the survey:

4.5.1.1 Individual Health Factors

What health information about the respondents do we have? Is it sufficient enough to process and find patterns or come up with recommendations or conclusion?

4.5.1.2 Individual Non-Health Factors

Does the dataset contain non-health factors such as demographic information, sex, age, marital status, and geographic information?

4.5.1.3 Community Health Profile

How the geography of a respondent can affect their responses? What is the overall health facilities and accessibility in that particular geographical setting? How the community makes its living? How life goes on every day in that community? What factors can make or break the health landscape of that community?

4.5.1.4 Community Non-Health Profile

How many dimensions are used in our survey besides health indicators?

4.5.1.5 Economic/Cost Factors

Do we have respondent's co-morbid conditions? How he/she is making his living? How much? What economical and cost factors influence his/her decision to get quality healthcare?

4.5.1.6 Educational Factors

What are the educational backgrounds and literacy levels of survey respondents? Is there any help needed to explain the questions? Which languages have been used?

4.5.1.7 Distance Factors

How the distance (travelling time) to health facilities is affecting the overall healthcare and impacting the outcome of intervention? How much travelling is needed for each survey respondent? How much do they need to travel for other necessities of life?

4.5.1.8 Surveyors Factors

How is the issue of bias handled by surveyors? Are they educated enough to guide the respondents? How many surveyors have to be employed to collect the dataset? Who is responsible for collecting a particular data row?

4.5.1.9 Socio-Political Factors

What role do the socio-political factors play in one's life for making health decisions? How the incumbent government or ruling party affects the health scenario in the community?

4.5.1.10 Availability of Health Services

How many healthcare facilities are available where survey participants can reach? What is the level of quality there? What types of services are present? How does the referral to larger and more modern facilities work? How long does it take to get there from the primary health facility?

4.5.2 Pilot Study

Based on the above mentioned topics, survey questionnaire was distributed among 300 families living in the proximity of Haripur district as a formal pilot study. Initially, our aim was to refine our survey questionnaire by collecting the data from a set of people to understand the ground realities and the problems that these people face every day. Some families came to the hospital or local clinics to fill the questionnaire while others were contacted at their homes. The questionnaire was given to the household heads aged 21 years or over, with one or more family members living with them, who were mentally stable and willing to participate in the survey. Furthermore, there was no race, religion, and gender discrimination during the survey activity.

During this pilot survey, we also convinced the village elders to actively participate in the survey and to provide us with a good insight about available health facilities in their villages. With the active participation of the village elders, we met a number of families and explained to them the purpose of this survey and the changes that it could bring to their daily lives. As a result, all the families who participated in this survey were provided with free medical services in order to retain their interest and to motivate other families living in the same region. Furthermore, in order to maintain the confidentiality and integrity of the households, their names were kept secret and were masked in the dataset.

As a result of pilot study, the following information was gathered which helped us in re-shaping/re-designing the survey form:

- Most of the participants have difficulty getting clean water or do not have a proper toilet facility. Almost, all are single with at least two adults living with them.
- Students from participating families have to travel for at least 20 minutes to get to school.
- Most families fall short of money if anyone in the family falls ill or has an injury.
- Stone and mortar is mostly used as the basic construction material of external walls while ceilings are mostly made of thick wood.
- Waste food, water and garbage are mostly disposed of near homes.
- During the whole year, people have to rely on an irrigation canal for a water source.
- Participants rely mostly on land and use it for agriculture or livestock.
- Most common hardships faced are loss of job or losing a house.
- Houses with toilet facilities have a very poor sewerage system.

Furthermore, based on the feedback, we added some other information in the survey that affects the overall healthcare system in Pakistan, including: distance to the health units, number of basic health units, costs associated with travelling, accessible routes to the health facilities, vaccination, transportation, sewerage system, awareness, and water resources, just to name a few.

Here, it is worth mentioning that similar questionnaires can be found in studies carried out by PPAF and Microfinance Risk Management L.L.C. (GoP, 2012, 2015; Schreiner, 2010). However, PSP (Schreiner, 2010) focusses only on the poverty dimension and the PPAF questionnaire is intended to gauge the impact of PPAF activities for poverty alleviation (GoP, 2012, 2015). In the proposed research, the questionnaire is designed particularly for this research and is more comprehensive in terms of eliciting poverty indicators and their relevance to health impact parameters.

4.5.3 Data Collection

At the end of the pilot study, all the changes were incorporated in the questionnaire. During the pilot study, it was observed that on average it takes two days per household to complete and return the filled survey form. Furthermore, the process of digitization and validation of the collected data is time consuming, tedious, and error prone.

To cater for the time, cost, and errors associated with the process of digitization, the mobile phone in addition to paper-based questionnaire was adopted for data collection. In this regard, an Android based mobile application "Survey on Tap" has been developed, which not only helped in conducting the surveys but can also helped to monitor large-scale health survey programs. Furthermore, its bi-lingual support, i.e., Urdu and English, and its ability to work in both on-line and off-line mode makes it unique to its competitors. Fig. 4.4 shows some of the screen-shots of the developed mobile application which were originally designed in the national language of Pakistan, i.e., Urdu. One such application is "QuickTapSurvey" application which allows its users to create their own survey forms. However, it maintains the database locally, i.e., it is not connected to a central database which means that all the information is on the mobile phone or tablet. In case of mobile phone theft, all the personal information of the responders of the survey will be compromised.

In contrast to “QuickTapSurvey”, our application pushes the data to the central site once connectivity is being provided. Furthermore, it uses GPS services to determine the location and Wi-Fi/mobile data connection to push the information to central database where the collected data can be monitored and visualized.

Apart from the mobile application, it has a synchronised web-portal that allows its administrator to visualize and monitor the data collection activity in real-time. Furthermore, it supports role based privileges for both the responders and healthcare professionals.

Here, it is worth mentioning that on average, it took 37 minutes by the household to complete the survey form and submit it by using the proposed application. Apart from the quantitative analysis of the collected data, subjective evaluation has also been performed regarding the usage of mobile phone for the data collection activity. For this purpose, four different evaluation criteria have been adopted (Palazuelos et al., 2013), namely: comfort, acceptability, preference, and accuracy. Generally, most of the responders were comfortable and satisfied with the user interface. According to them, usage of mobile phones for the collection of health data is self-sufficient, saves time, and is easy to use compared to conventional paper-based forms. However, a few of the responders reported that due to a shortage of electricity, battery time, and small screens, it was not convenient to fill the survey forms using mobile phones.

In total, the BMWT dataset comprises of 47 features collected from 1000 households that plays a vital role in the healthcare system of Pakistan as show in Appendix E.1. Appendix E.2 shows the scaling of the BMWT features. Here, it is worth mentioning that in addition to features shown in Appendix E.1, monetary values were also noted. These monetary values include the transportation cost from the village to nearest health facility. On average, each family has to spend 3634 PKR to just reach the nearest government health units as shown in Appendix F.

Moreover, due to limited health facilities at the nearest health units, patients have to travel 38 kilometres (round trip) on average to the main hospital. This round trip costs them 9186 PKR on average. Due to the poor road condition and shortage of transport, it

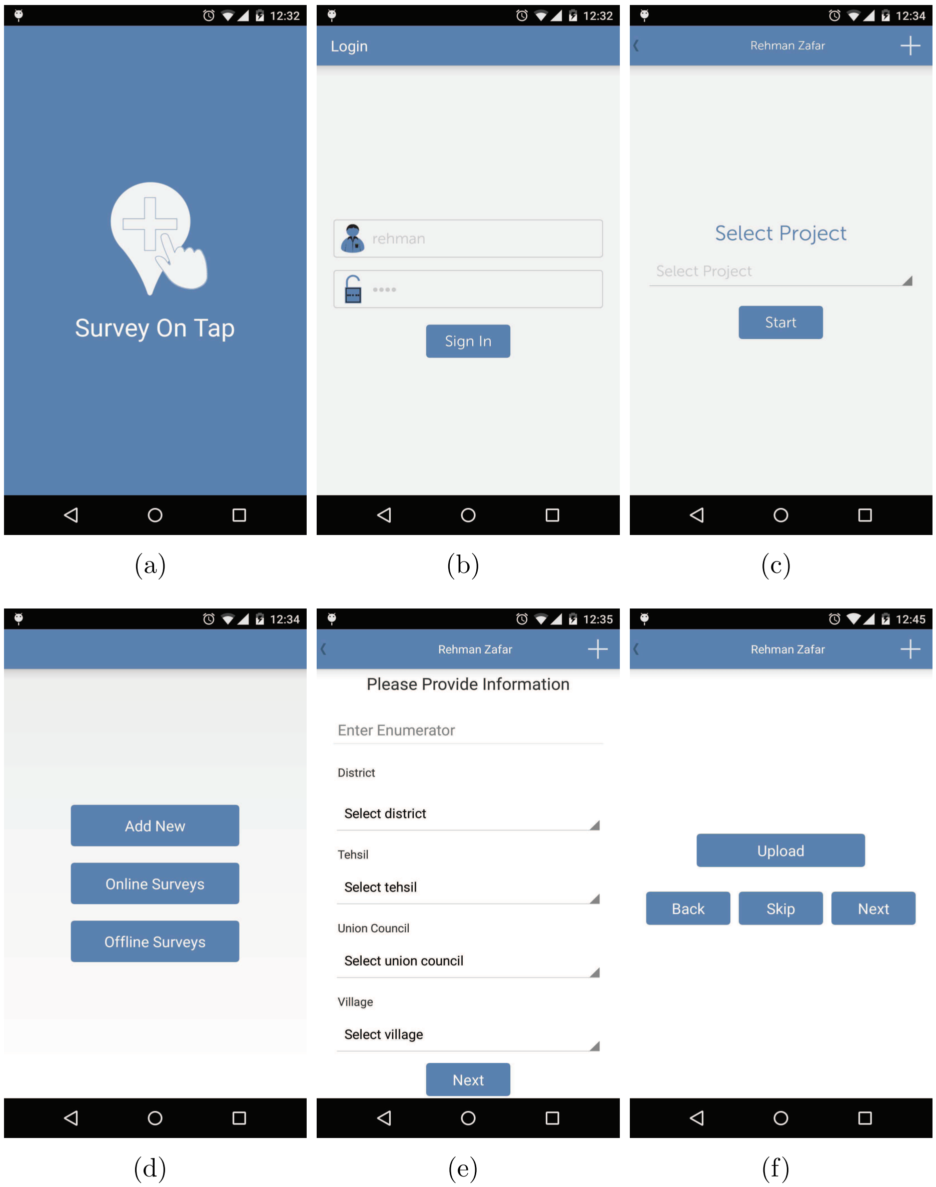


Figure 4.4: Survey on Tap. a) Welcome screen. b) Sign-in screen. c) Project selection screen. d) List of surveys done. e) Survey form screen. b) Successful message on completion.

takes from a few hours to one day to travel 38 kilometres. The high cost of travelling in comparison to the monthly income of each household is one of the major causes of health-shock in district Haripur where more than 42% of the population of the district is below the poverty line and 31% of the population is in the middle-income group that is living hand to mouth. Moreover, the unemployment rate in district Haripur is almost 30% (GoP, 2010).

4.6 Summary

Currently, there is no publicly available dataset that can help to understand and monitor health-shock in Pakistan. Such surveys and datasets can be helpful to government—who would use this dataset and resulting analysis to form policies, for general practitioners and NGOs, in order to start community based health programs. Therefore the dataset was collected with the support of BMWT. In total, BMWT dataset comprises of 47 features collected from 1000 households belonging to 29 villages in rural areas of Pakistan.

The purpose of this research was to develop the first contextaware health dataset that is purely based on the socio-economic, cultural, and geographic norms of Pakistan. Such datasets will help to understand the relationships between socio-economic, demographic, and geographical conditions and their impact on health. Moreover, one of the problems associated with big data could be its storage and retrieval. Therefore, a cloud based architecture is suggested to collect the data via mobile devices and store in the cloud.

The next chapter will present the data analysis, visualization, statistical and machine learning approaches used to analyze and visualize the collected dataset. Furthermore, cloud computing as an enabling technology is used in this research in order to deal with large-scale health informatics dataset.

Chapter 5

Cloud Enabled Data Analytics and Visualization Framework

5.1 Introduction

The previous chapter has discussed how the data was collected; and what protocols and procedures were applied to collect the dataset from 1000 households. This chapter analyzes the dataset using a statistical approach. Furthermore, in order to cope with a large scale health informatics dataset, this chapter proposes the use of cloud enabled framework for data analytics and visualization. The proposed framework is developed using cloud computing services based on Amazon web services (AWS) integrated with a geographical information systems (GIS) to facilitate big data capture, storage, indexing, and visualization of data through smart devices for different stakeholders.

The schema of this chapter is as follows: Section 5.2 presents the proposed cloud enabled framework for data analysis and visualization; Section 5.3 discusses the results of analyzing various variables in different villages of District Haripur such as monthly income, sources of income, travelling time and cost to BHUs from villages, debt situation, food disposal, land ownership, major and minor diseases and their correlations; Section 5.4 discusses the mapping of data onto the "Iron Triangle model of Healthcare"; and Section 5.5 provides the summary of this discourse.

5.2 Cloud Computing and Healthcare

In order to understand health-shock and the causes , there is a dire need to develop a framework that can help to collect, analyze and visualize the reasons behind health-shock in addition to producing mitigation policies at a national level (Jafar et al., 2013; Nishtar, 2011; Nishtar et al., 2003). However, such large-scale data collection, analysis, and visualization needs a large amount of financial and physical resources in addition to computation and storage. Here, a cloud computing enabled framework using mobile devices can play a vital role by reducing the time and cost associated with data collection, analysis, and visualization (Andreu-Perez et al., 2015; Manekar and Pradeepini, 2015; Rudan et al., 2005; Samuel et al., 2013). Furthermore, with the reduction in prices of mobile phones and almost universal coverage, it can make healthcare more accessible at a lower cost by taking geographical distances out of the equation (Herb et al., 2012; Velthoven et al., 2013). Moreover, such large-scale data analysis and visualization will help to modify and understand the iron triangle under the socio-economic, geographical, and cultural norms of any country.

In this regard, a cloud enabled framework has been proposed in which all the high performance and computing requirements of a GIS system, including data gathering and cleaning, in addition to data analysis and visualization at a large-scale have been fulfilled by cloud computing. Here, GIS helps to provide us with highly accurate and interactive maps based on the gathered data. Furthermore, we have also developed a mobile application that enables the healthcare professionals to create, and deploy health surveys. As a whole, all data will be collected from mobile applications by using web-services. Moreover, the cloud enabled framework will allow us to apply the machine learning algorithms at high speed. Furthermore, data will be visualized on maps using a geographical information system.

5.2.1 Amazon Web Services as Cloud Platform

Here, we have used Amazon Web Services (AWS) as a cloud platform. AWS contain an array of data centers and services, and host around 27% of global cloud computing market. One in every three website visits globally are hosted on AWS. Furthermore, cloud computing has reached 40 Billion USD market today, and is projected to grow to 241 Billion USD in the next 10 years by utilizing the robust, scalable and affordable infrastructure of AWS.

AWS provides turnkey solutions for cloud computing and has dozens of ready-to-use services for all components of cloud infrastructure. Its most famous service called EC2 (Elastic Computing Cloud) lets the user select storage and computing resources on the go. User can purchase the computing instances in three ways: on-demand, reserved, and spot instances. On-demand as the name suggests, let the user pay per use without any commitments. On the other hand, reserved instances can be purchased ahead of time for a discounted hourly rate whereas the spot instances incentivize companies to use AWS resources during off-peak hours or times of under-utilization for a better price. The price in this instance fluctuates based on demand and availability of resources.

AWS is readily distinguished from other vendors in the traditional IT computing landscape because of its openness, flexibility, cost-effectiveness, scalability, elasticity, security, deployment speed, and performance. It allows users to use any programming model, databases, operating systems or software architecture of their choice with mix-and-match architectures on the go and makes it the most flexible cloud platform in the world. It is economical in a way as users only pay for what resources they have used. Users can also add and delete AWS resources for their applications that make it very elastic for deployment. AWS offer world class security encryption for both in-flight IOPs and at-rest IOPs. It has a complete suite of client certification and accreditation, with identity and access management. AWS has reduced the server deployment time from days to few minutes. Its S3, Elastic Block Storage (EBS) coupled with backup of Amazon Glacier, makes it the best storage choice in cloud industry. AWS has an incredible bandwidth with reliability of 99.999999999% that makes it the best performing cloud platform in the world.

Fig. 5.1 presents the complete technology landscape for using cloud computing to perform data analytics and visualization. The first layer is of databases, one can use any flavour based on application requirements. Since most of the analytical process go well with column based or structured databases, we have recommended HP Vertica, Oracle, and SciDB. The second layer of cloud computing contains the basic parts from AWS, i.e., EC2 for computing, S3 for storage, VPC for the virtual cloud, and Glacier for automated backup. Layer 3 takes care of hosting services by Route 53 from Amazon and IIS. Layer 4 recommends our choice of operating systems for such application. Layer 5 gives a snapshot of available tools and utilities that go by with such applications and the last layer chalks down the quality assurance requirements for data analytics applications in the cloud.

However, in our implementation, we have used the technology stack as shown in Fig. 5.2.

5.3 Data Analysis and Visualization

Data analysis and visualization is vital to understand the hidden patterns in data that, when analyzed can lead to policies and recommendations for better planning, cost efficiency, and thus improved quality of life for the target population. In this regard, we created a tool formally known as "HexChange" in dot Net for data analysis and visualization. Here, we have implemented various methods such as decision trees (Quinlan, 1987; Strobl et al., 2009), fuzzy c-means clustering (Dunn, 1973), and fuzzy linguistic summarization (Doctor and Iqbal, 2012; Meunier and Moyse, 2012; Niewiadomski, 2008) on cloud using AWS and the proposed technology stack.

Here, it is worth mentioning that before data analysis and visualization, the data instances with values out of the predefined data ranges were removed in order to decrease the noise level of the dataset. Moreover, after the removal of outliers, the data categories of categorical variables were ranked by giving each category an ordinal number based on expert opinion. Finally, in order to unify the range of the data for the all considered variables so as to take values between 0 and 1, the data was normalised by dividing each data instance by the maximum value of the variable which it represents.

As discussed above, in total, BMWT dataset contains 1,000 households from 29 villages of 5 different union councils of district Haripur as shown in Fig. 5.3.

Figure 5.1: Technology Stack for Big Data Cloud Computing, Data Analysis and Visualization.

Figure 5.2: Used Technology Stack for Big Data Cloud Computing, Data Analysis and Visualization.

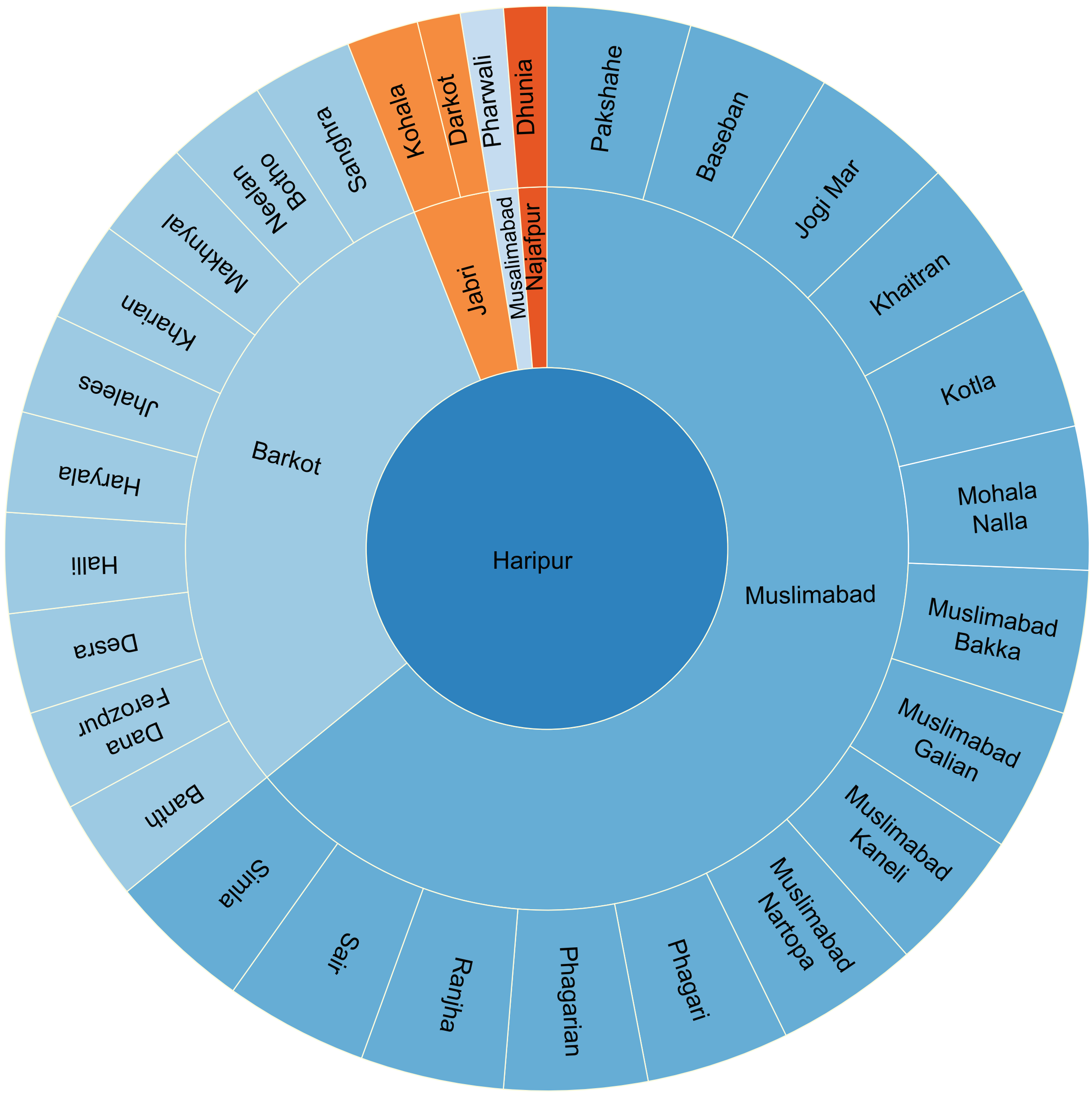


Figure 5.3: Villages that participated in the survey. Inner most circle represents the district Haripur whereas the outer circle represents the union councils and the outermost circle represents their corresponding villages.

5.3.1 Data Normalisation

In district Haripur, more than 42% of the population are below the poverty line; whereas, 31% of the population are in the middle-income group. Here, poverty in addition to the limited commute resources, is affecting the education in the district.

Fig. 5.4 shows the population split and their reach to schools. As shown in Fig. 5.4, Muslimabad is the largest with 47% population, Barkot has 33.8%, Najafpur 10.6%, Jabri 6.3% and Musalimabad 15%. The residents of Musalimabad have to travel larger distances to reach schools. Schools are most accessible for residents of Jabri. The interquartile range is within 7 KMs for all 5 Union Councils. Similarly, the residents of Barkot and Muslimabad

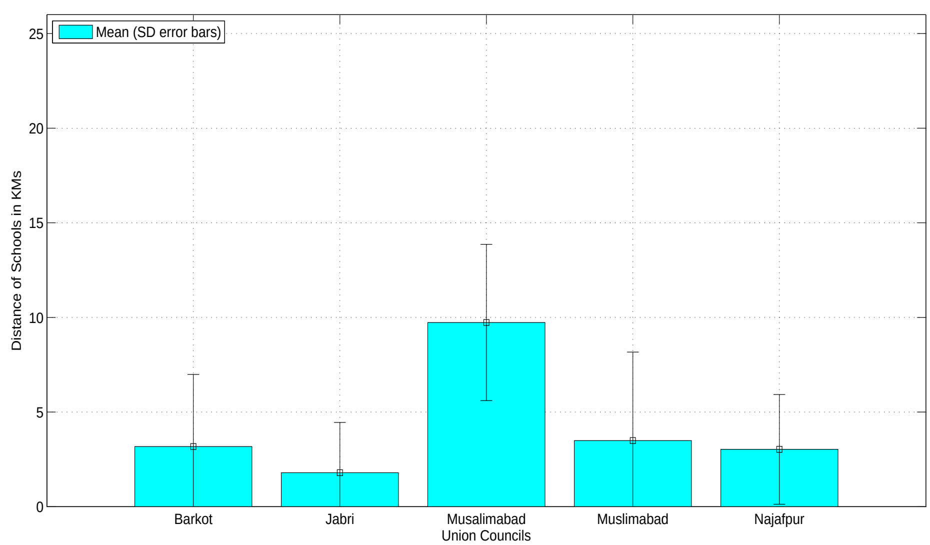


Figure 5.4: Population Split and Reach to Schools.

have to travel varied distances to reach a BHU as shown in Fig. 5.5. BHUs are most accessible for residents of Jabri and Najafpur.

Appendix G.1 – Appendix G.5 presents the thematic map of distance to BHUs and main hospital from the villages. Here, the minimum travelling cost from village to BHU is approximately PKR 1,200 to PKR 8,000. In contrast to travelling cost, 42% of district Haripur is below poverty line, where 13.3% of the population have a monthly income of less

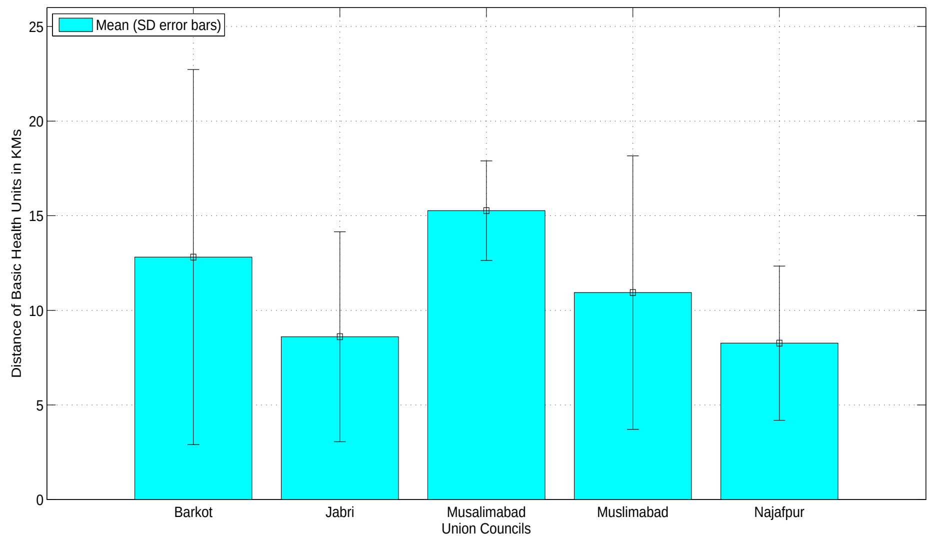


Figure 5.5: Distance to Basic Health Units (BHUs).

than PKR 5,000. Moreover, in case of emergencies or major operations, people of district Haripur have to travel to the main district hospital where the travelling cost ranges from PKR 6,200 to PKR 1,4000. For the majority of people, the most obvious and first choice to solve the above mentioned problems is to borrow money from their relatives and friends as shown in Fig. 5.6. Even in the case of personal violence problems, residents borrow money to resolve the issues, which indicates that most of the personal violence issues are related to money. Hence, these problems force people to rely on debts, which acts as a major cause of health-shock as shown in Fig. 5.7. It also shows that the majority of people are in debt from 20,000 PKR to 500,000 PKR and above.

Furthermore, the majority of houses have stone-and-mortar walls along with thick wood ceiling. Metal sheeting is mainly found in Barkot (28% residents) and Muslimabad (13% residents). Houses will get damaged due to the harsh weather with equal chances of major and minor damages. Most of the reconstruction is usually completed within 2 years of damage. No significant relation was found between house strength and ownership of the house. The gender distribution of the Haripur Villagers shows that Jabri has reported 100% Males and Barkot, Muslimabad, Najafpur has a majority of males. Only Musalimabad

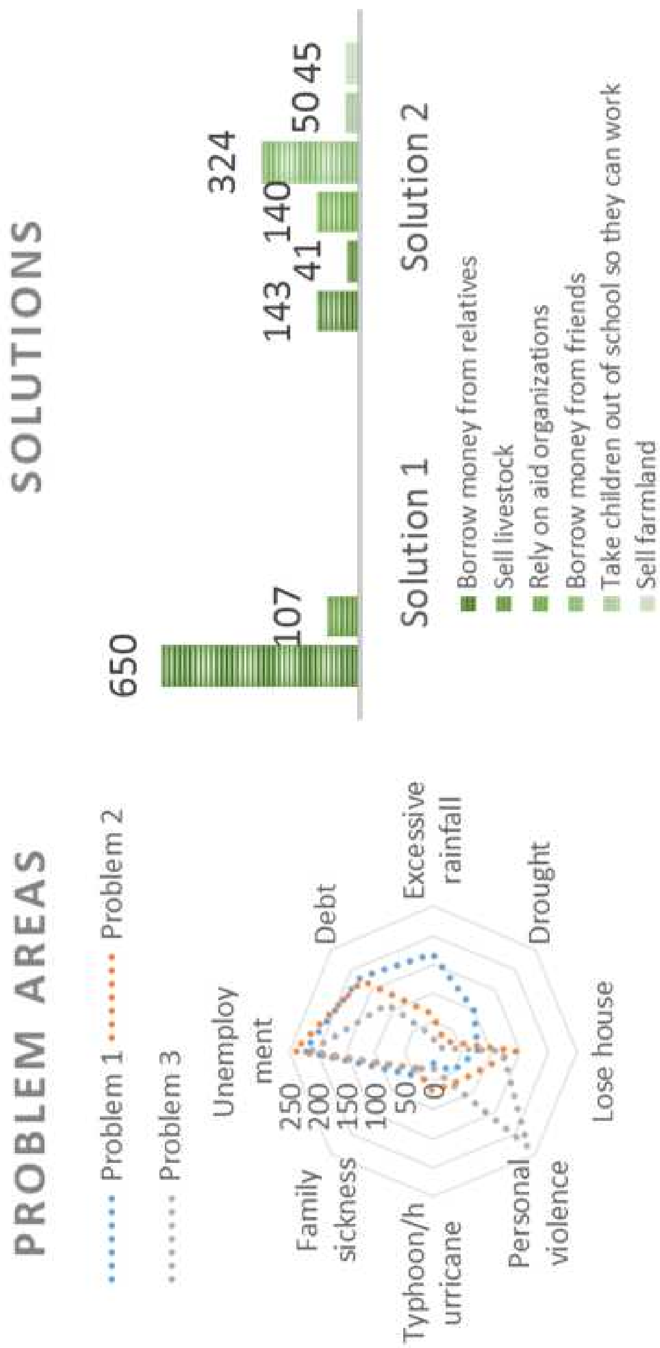


Figure 5.6: Spiral of Problems and Debts.

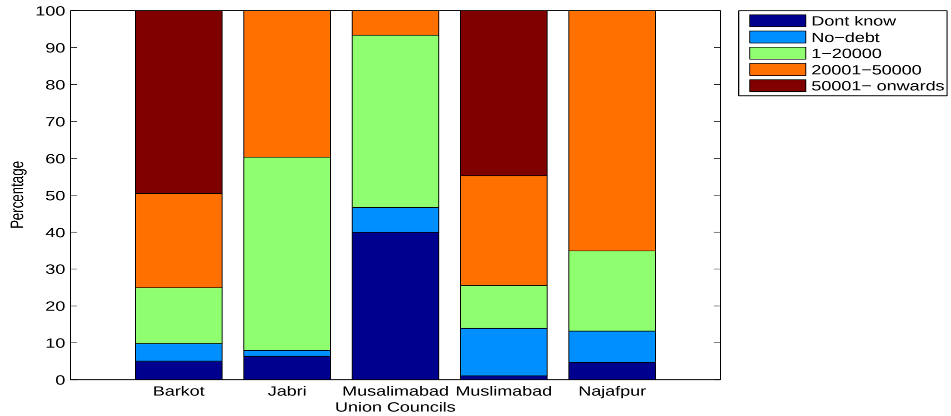


Figure 5.7: Debt Scenario in District Haripur.

reported a majority of females, i.e., 66.67%. Fig. 5.8 shows the number of permanent adult members of the household. The majority of households have 2 to 4 adults. The household size is right-skewed and there are a small number of families with much larger number of adults.

Here, it is worth mentioning that in the BMWT dataset, there is a positive relationship between distance to basic health units (BHUs), percentage of debts, toilet facilities, and frequency of major illnesses. Due to debts and distance to BHUs, minor illness turns into major illness. Similarly, dependency ratio, which is defined as the age-population ratio of those typically not in the labour force (i.e. the dependent part) and those typically in the labour force (i.e. the productive part), ranges from 3.4 to 4.9 for villages of district Haripur. Fig. 5.9 shows earning adults, total adults and the dependency ratio. Barkot has the highest dependency ratio, while Jabri has the lowest.

In Haripur, water is disposed mostly within 75 meters of home or piped down the drain. Very little is used for irrigation. Garbage is discarded within 75 meters of the house or else burnt. Food is fed to the livestock if not disposed within 75 meters as shown in Fig. 5.10.

Fig. 5.11(a) presents the situation concerning land ownership. Muslimabad has the highest number of private owners. However, a majority of respondents refused to share the acreage details of their land. Furthermore, no relation between size of land ownership and

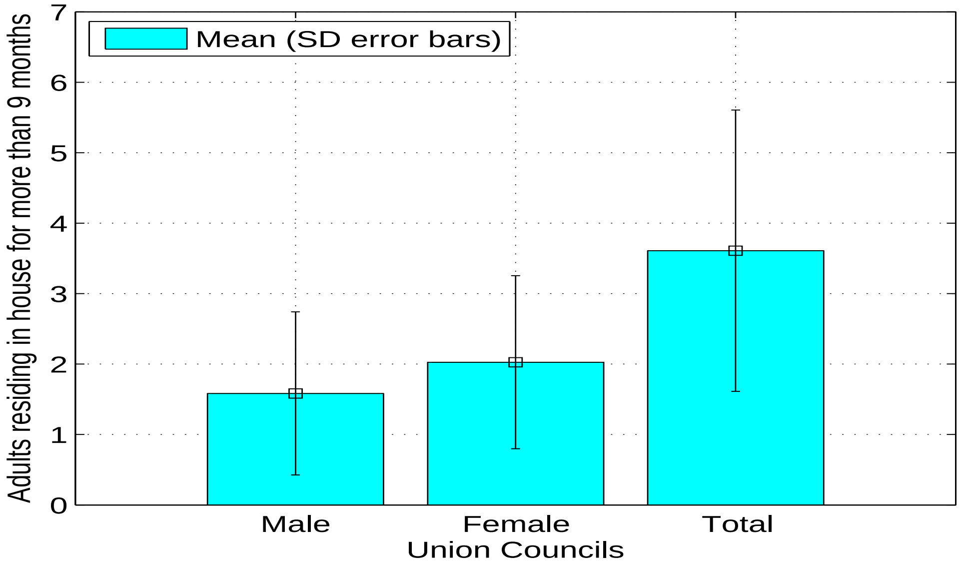


Figure 5.8: Adults Living in the Household.

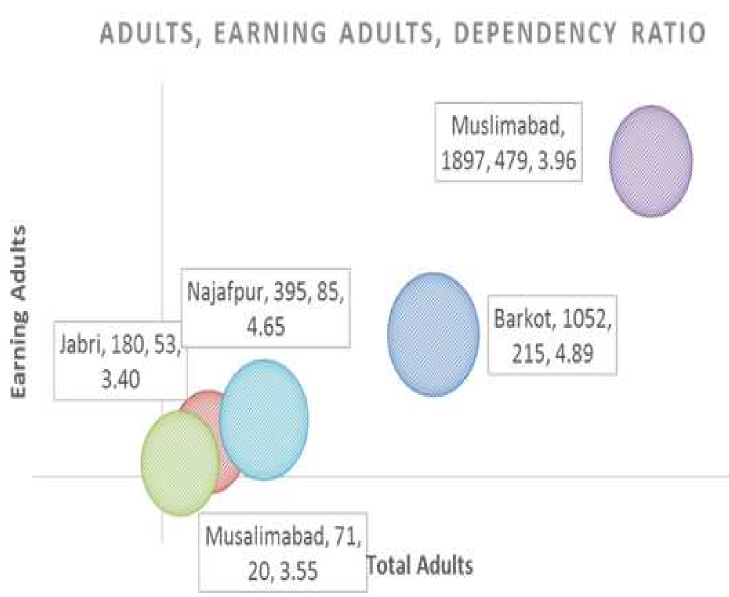


Figure 5.9: Dependency Ratio.

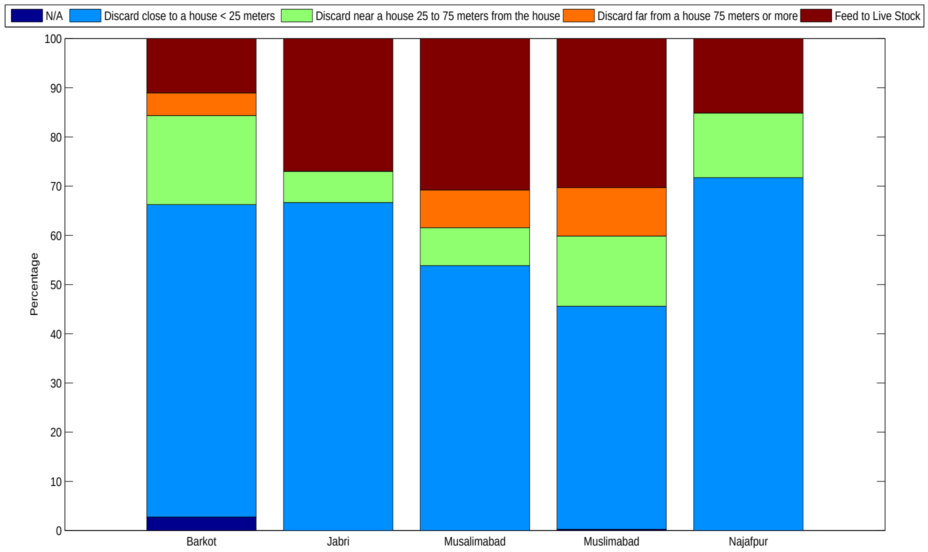
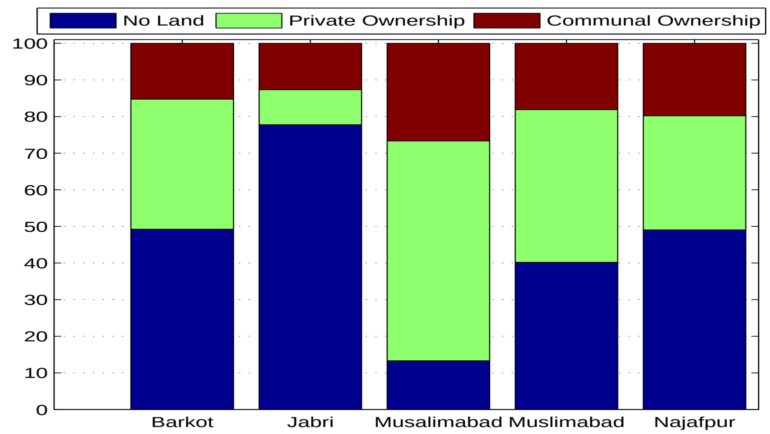


Figure 5.10: Food Disposal.

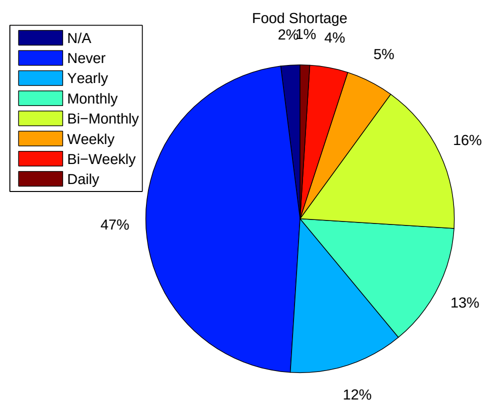
type of land ownership was found during the survey. Similarly, Fig. 5.11(b) presents the food shortage. Based on BMWT survey, 47% households have never faced a food shortage, while 4% face bi-weekly food shortage.

Furthermore, it was shocking that houses with toilet facilities have a higher rate of diseases in comparison to houses with no toilet facility. One of the main reasons was poor access to water resources and poor sewerage system. Here, it is worth mentioning that the time required for one person of a household to collect water for one day usage is almost 4 hours as in some cases, it takes a women 1.5 to 2 hours to reach the source. The same amount of time is required to carry that water back home. Especially, in the case of a family with two to three children, it requires more than 5 or 6 buckets of water at least for a day. Fig. 5.12 shows the relation between the time required for collecting water and the sewerage system deployed.

Similar to this, a relationship has been observed between the frequency of minor/major diseases and toilet facilities as shown in Fig. 5.13. Here, minor disease is defined as any normal disease or injury which doesn't require bed rest whereas any disease or injury which require 2 or more days of bed rest or hospital admission, including disability, is represented



(a)

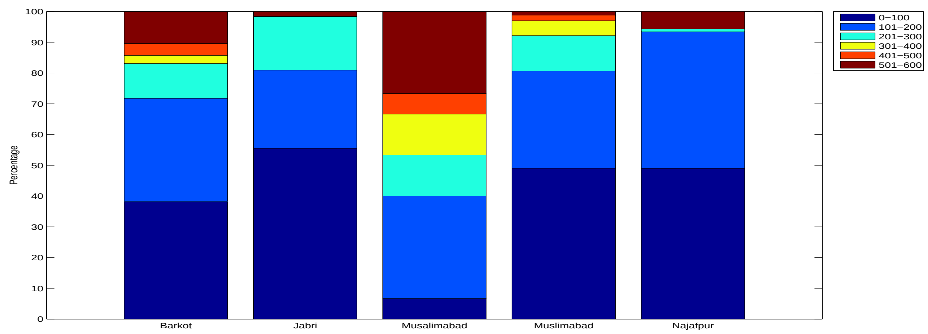


(b)

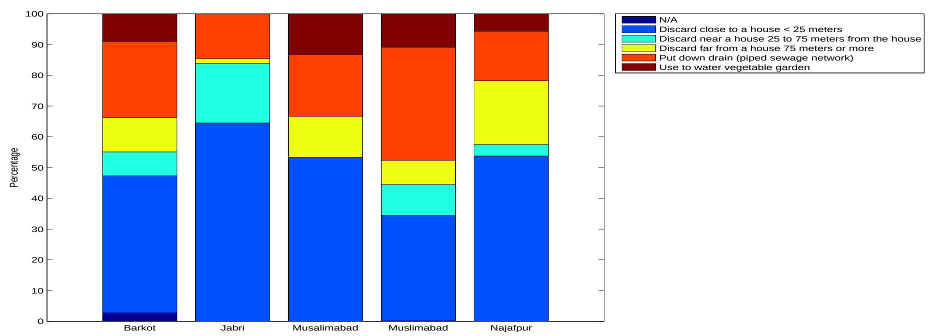
Figure 5.11: a) Land ownership. b) Food shortage.

by major disease.

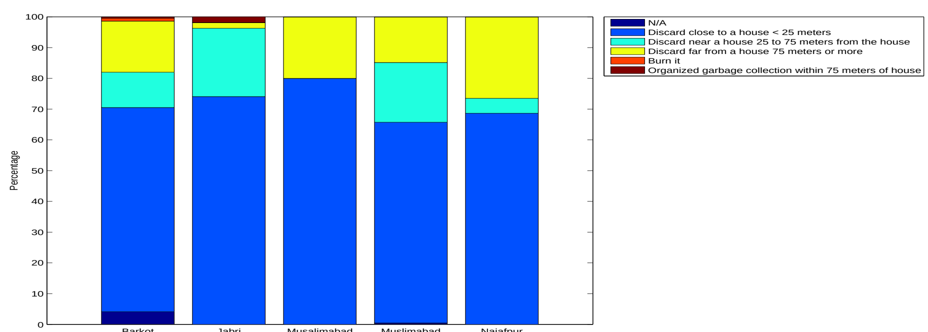
During the survey, most reported problems were unemployment, debt, climate, excessive rainfall, and personal violence. In district Haripur, climate is hot in summer and cold in winter. Spring and autumn are the transitional seasons between summer and winter. December to February are the coldest months where temperature falls down to $1\,^{\circ}\mathrm{C} - 3\,^{\circ}\mathrm{C}$ as shown in Table 5.1. Moreover, a gradual change in the climate has been observed by the local communities in the past twenty to fifty years. The temperature and rainfall schedule is becoming erratic year by year, causing adverse impact on farming and other routines of the



(a)

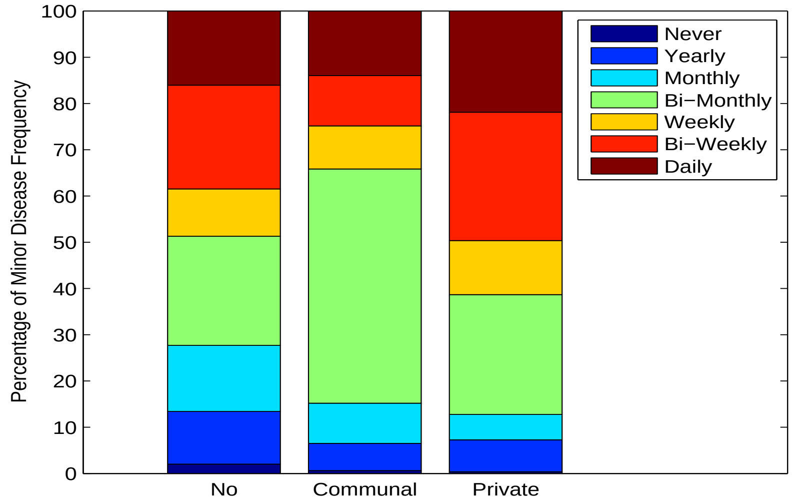


(b)

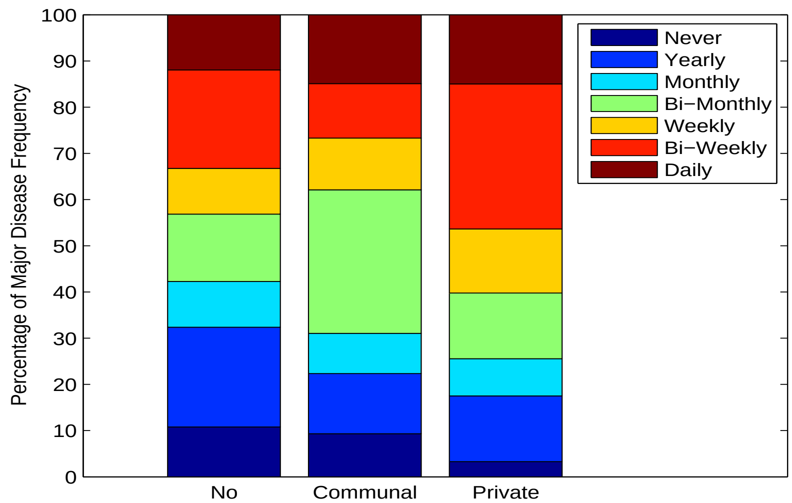


(c)

Figure 5.12: a) Time required for collection water for one day in minutes. b) Water Disposal. c) Garbage disposal.



(a)



(b)

Figure 5.13: (a) Frequency of minor disease versus toilet facility. Here, communal represents a toilet shared by more than 3 people. b) Frequency of major disease versus toilet facility.

Table 5.1: Month Wise Mean Temperature, Precipitation and Relative Humidity, District Haripur (GoP, 2010).

Month	**Maximum Temp. in °C**	**Minimum Temp in °C**	**Precipitation (millimetres)**	**Relative Humidity (%)**
January	12.6	1.8	64.8	59.1
February	13.4	2.9	113.6	61
March	17.8	6.9	142.3	57
April	23.3	11.4	111.8	51.4
May	28.2	15.5	81.6	42
June	32.4	19.7	85.3	41.2
July	29.6	20.1	258.3	66.7
August	28.2	19.3	261.3	47.7
September	27.8	16.8	96.9	62.3
October	24.9	12	56.9	51.3
November	20.1	7.2	31.9	49.4
December	15	3.4	61.5	55.9
Annual	22.8	11.4	1,366.20	56

community, e.g., due to severe drought. As a result, the farmers are giving up crops, which need more water and work load of women has increased, e.g., in the form of water collection. Erratic rains lead to floods in rain fed stream that cause damages to agriculture lands and other infrastructure, while it also degrades range lands perpetually.

5.4 Mapping Data Onto Iron Triangle

As discussed in section 2.2, healthcare systems are evaluated based on three main factors: quality, cost, and accessibility to healthcare, known as "The Iron Triangle Model of Healthcare". However, these factors highly depend on the socio-economic, geographical and cultural norms (Mahmud et al., 2014).

In order to understand the iron triangle model, we performed data analysis and visualization on the collected dataset. By doing so, we were able to understand and map the iron triangle model under the socio-economic, geographical, and cultural norms, and factors that lead to health-shock in Pakistan. Moreover, in order to develop a framework for health-shock

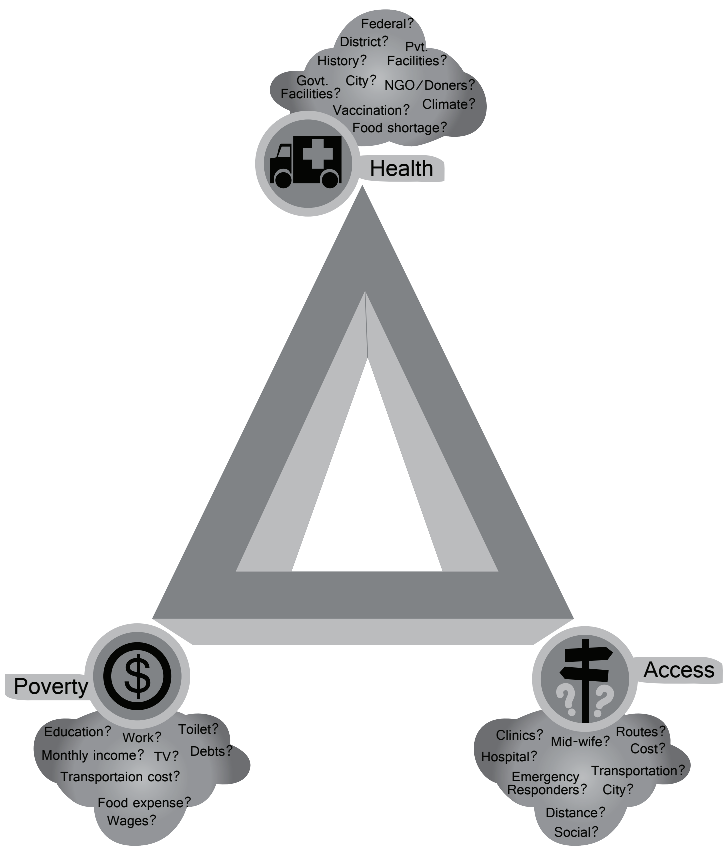


Figure 5.14: Factors affecting the Healthcare system of Pakistan.

prediction, it is extremely important to identify the social factors operating in a society like the socio-economic indicators for the household.

Furthermore, to estimate variables like consumption and expenditure including food expenditure, health expenditure, etc., a household asset index (based on domestic animal ownership) was used as indicative of a household's permanent income. Details of these variables are given below:

5.4.1 Control Variables

Social and economic conditions of the area always play a vital role in the wellness of the society. Especially, when it comes to healthcare, indicators such as monthly income, health expenditure, food expenditure, transportation cost, climate, frequency of disease, and debts, should be measured and mapped onto the iron triangle. In this regard, we have mapped these factors onto the iron triangle as shown in Fig. 5.14.

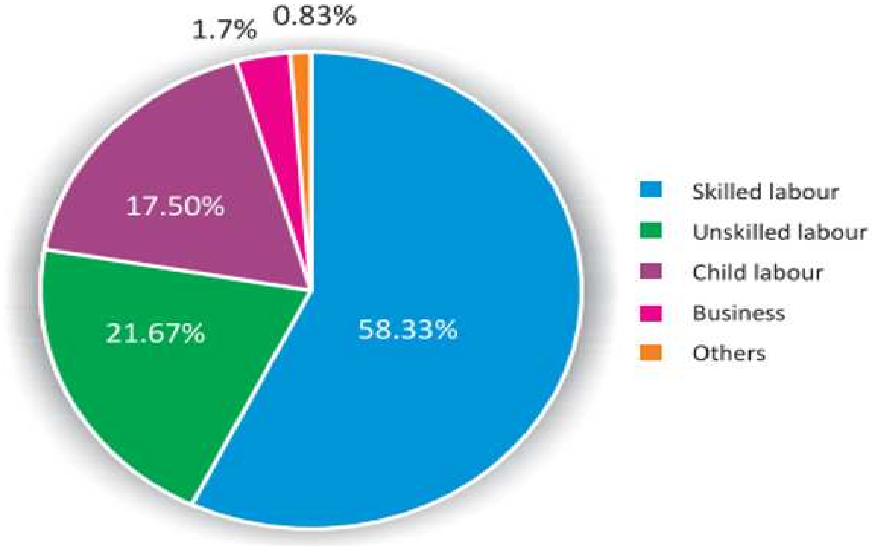
5.4.1.1 Monthly Income

Monthly income of the family depends on the number of persons who are involved in skilled labour, unskilled labour, and child labour. Fig. 5.15(a) shows the percentage-wise breakup of the source of income in the area. Fig. 5.15(b) highlights that 13.33% households enjoy total monthly earnings of PKR 15,000 per month, 60% are earning less than PKR 18,000 per month and approximately 32% population are earning below PKR 15,000. The graph shows the minimum monthly income as PKR 5,000 and maximum as PKR 35,000 only. We have distributed the income of the sample population according to given percentages.

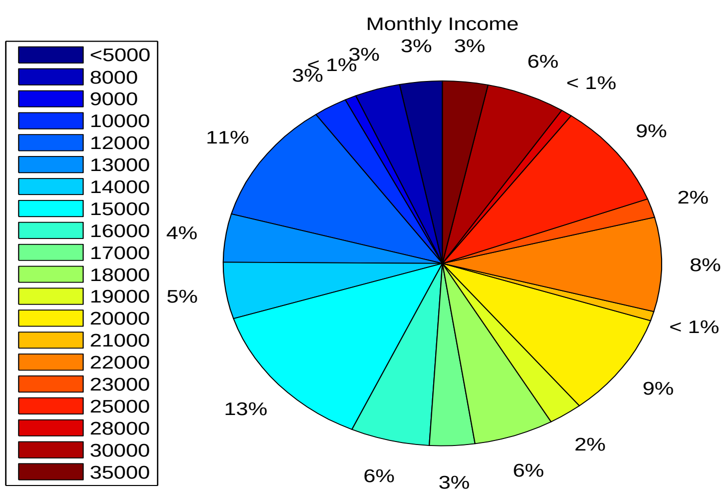
5.4.1.2 Food Expenditure

According to a report by the US Department of Agriculture's Economic Research Service, people in Pakistan spend more income on food than people in any other country. An average Pakistani spends 47.7% of the household budget on food consumed at home.

District Haripur is comprised of 88% rural population, so we have considered the rural



(a)



(b)

Figure 5.15: (a) Sources of income. b) Monthly income of Households in District Haripur.

class expenditure on food with the average and high-income groups as shown in Table 5.2. The people from the average income group spend 63% on all kind of food and the people from the high-income group spend 44%. The household size also affects the expenditure on food.

Table 5.2: Expenditure Elasticities, KPK Rural Urban (Ahmad et al., 2015). Here, WF represents Wheat Flour, RF represents Rice Flour and MP represents Milk Product.

	Rural Income Groups					**Urban Income Groups**				
	Low IG		**Mean**	**High IG**		**Low IG**		**Mean**	**High IG**	
Commodity Groups	**IG 1**	**IG 2**	**IG 3**	**IG 4**	**IG 5**	**IG 1**	**IG 2**	**IG 3**	**IG 4**	**IG 5**
Wheat & WF	0.36	0.25	0.41	0.19	0.11	0.45	0.2	0.19	0.25	0.14
Rice & RF	2.89	1.71	1.25	1.22	0.79	2.6	2.23	0.75	1.04	0.46
Pulses	0.22	0.15	0.15	0.14	0.13	0.49	0.26	0.26	0.29	0.19
Milk & MP	0.93	0.82	0.7	0.65	0.47	1.38	0.65	0.57	0.93	0.34
Meat & Fish	3.99	3	2.36	2.44	1.23	2.61	1.13	0.84	1.59	0.45
Poultry	2.74	2.14	1.57	1.54	0.88	4.36	2.54	1.12	2.79	0.54
Fruits & D	5.82	2.34	1.8	1.76	1.02	2.15	1.37	0.79	1.67	0.42
Vegetables	0.98	0.59	0.57	0.54	0.44	0.58	0.55	0.32	0.48	0.21
Tea & Soft Drink	0.75	0.57	0.49	0.46	0.37	0.32	0.22	0.2	0.31	0.14
All Food	0.98	0.72	0.63	0.61	0.44	1.03	0.87	0.46	0.72	0.29

5.4.1.3 Health Expenditure

Appendix H represents the consulting fee for a doctor in various cities of Pakistan. Here, we have considered Peshawar as a baseline for Haripur district as it is the capital of Khyber Pakhtunkhwa. Furthermore, the compound annual growth rate (CAGR) formula has been applied to calculate the fee for the doctor for the recent year. The CAGR is 13.62% from 2008 to 2011. For 2015, the consulting fee for a doctor in Peshawar is approximately PKR 315. As we know, the monetary values of the rural area are 4 to 6 times less than the urban area of any location and as the Haripur District is comprised of 88% of rural locations, we have considered PKR 52.5 as the doctor fee in Haripur, including medicine charges as average. The number has been also confirmed by survey respondents from the BMWT hospital.

5.4.1.4 Transportation Cost to Hospital

In emergency, people use private transportation from house to health facility and back. The cost of a round trip has been considered as a journey from the village to hospital and back as shown in Table 5.4.

Table 5.3: Average transportation cost of round trip from village to main Hospital. Here, the travelling cost was calculated with the help of patients and local transporters.

Village	**Cost in PKR**	**Village**	**Cost in PKR**
Akhorra	9000	Banth	11000
Badhaar	7400	Hariala	8200
Bandi Kiala	6600	Kharian	7400
Serbaroot	13000	Neelaan	11000
Shah Kabal	12000	Kotla	12000
Tehal	7200	Makhrial	6600
Darkote	8000	Jandi	8600
Seri	6400	Desra	8200
Bandi Kiala	6200	Beesbaan	9400
Kohala Bala	7000	Baghpur Dheri	11400
Darkote	7400	Khoi Kamman	12600
Kohmal	8200	Bharrey	14000
Ramial	8600	Khoi Bagran	11400
Jalees	10000	Pakshahi	8600
Dhunian	9000		

Here, it is worth mentioning that villagers facing less travel cost are not falling ill frequently. Furthermore, if we predict severe disease frequency by transport cost then villagers with a low travel cost are mostly facing a yearly disease frequency and villagers with a high travel cost are facing a bi-weekly disease frequency as shown in Fig. 5.16. Fig. 5.16 shows that 56% of the villagers with a low travel cost are facing severe disease Yearly. It shows that they will not be experiencing health-shock. While 44% of the villagers with a high transport cost

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be viewed in the Lanchester Library Coventry University.

Figure 5.16: Relation between travelling cost and severe disease by using decision trees (Quinlan, 1987).

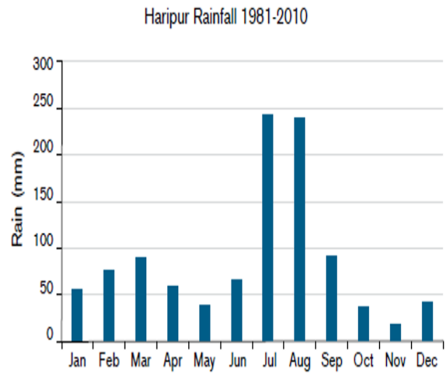
to the health unit are experiencing severe diseases bi-weekly, which clearly shows that they are having health-shock frequently.

5.4.1.5 Climate

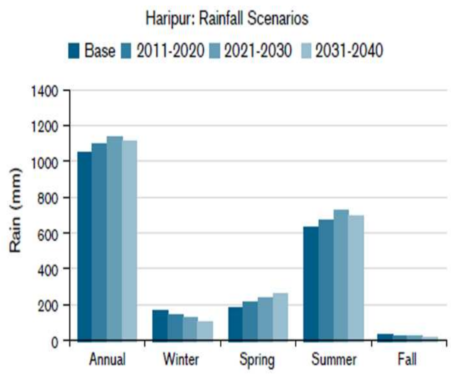
In Haripur, heavy rainfall is recorded in the months of July and August, which is more than 225 millimetres from the past 25 years. So, the overall rainfall is high in summer. Excessive rainfall, hurricane, strong wind, acid rain and flood are the common climate difficulties of the area. It is important to consider the current climate at the time of health-shock prediction. On average in the case of climate problems, 20% of their income is spent on the maintenance of household goods and farming. Fig. 5.17 presents the rainfall scenarios in the Haripur district.

5.4.1.6 Frequency of Disease

Income is the most important economic factor in the life of a villager. Here, we have tried to find a relation between income and the frequency of disease by using decision trees (Quinlan, 1987). From Fig. 5.18, it can be inferred that there is a 49% chance that villagers with low income will face a bi-weekly disease frequency and there is only a 22% chance that villagers with a high income will face a bi-weekly disease frequency. It shows that villagers with a



(a)



(b)

Figure 5.17: a) Month wise rainfall in Haripur district. b) Season wise rainfall in Haripur district.

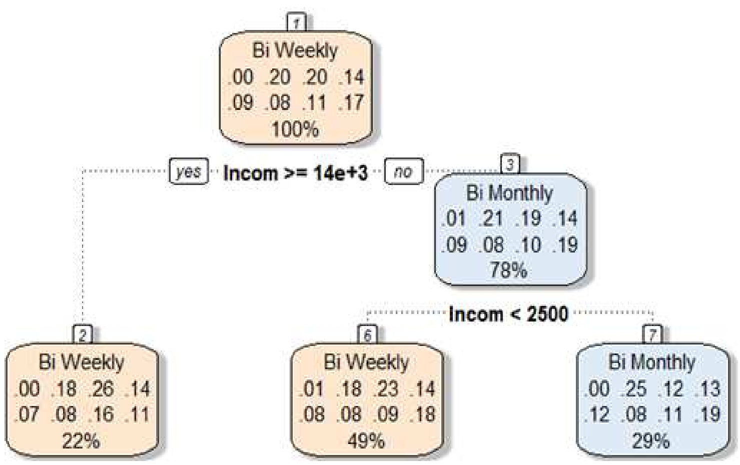


Figure 5.18: Relation between income and disease frequency.

low income are frequently facing diseases.

5.4.1.7 Debt Amount

Debt is a very crucial socio-economic variable for predicting health-shock. In rural area, most of the people are under a high amount of debt. According to survey data, 91% people are under the burden of debt and 49% of the population owe more than 50,000 PKR. Moreover, when someone gets sick and does not have the money to visit a doctor, the first choice is to get a loan and pay it back in instalments. Here, it is worth mentioning that the interest on these loans is as high as 100% but poor villagers have no other choice. The situation becomes even more complex in the case of a climate problem or any other negatively influencing variable.

5.5 Summary

Cloud computing technology provides a plethora of benefits which can be of great help in real time health-shock monitoring and data analysis. With the usage of mobile technology, large-scale health surveys can be conducted with minimal human and financial resources. Moreover, cloud computing enabled GIS applications in addition to data visualization can play a vital role, especially within Pakistan. In this regard, a cloud enabled framework for data collection, analysis, and visualization has been proposed and discussed in this chapter. Such a framework has helped us in modifying the "Iron Triangle model of Healthcare" according to the under the socio-economic, geographical, and cultural norms of Pakistan. Moreover, such a framework in addition to the dataset created will enable the healthcare professionals to collect, analyze, and visualize healthcare at a minimal cost with robust results. Furthermore, advantages of cloud computing over traditional hosting and cost effectiveness of AWS have also been discussed in this chapter.

The next chapter will present the proposed Fuzzy rule based summarization approach for health-shock prediction.

Chapter 6

Proposed Framework for Health–Shock Prediction

6.1 Introduction

The previous chapter has presented the analysis of the data captured from 1000 households and provided a means to map the dataset onto the Iron Triangle Model of Healthcare. In the previous chapter, the data has been analyzed using various statistical approaches. However, to predict heath-shock, this chapter presents a fuzzy rule summarization based predictive model. Here, the collected data has been used to generate a predictive model of health-shock using a fuzzy rule summarization technique, which can provide stakeholders with interpretable linguistic rules to explain the causal factors affecting health-shock. Here, it is worth mentioning that the evaluation of the proposed system in terms of the interpretability and classification accuracy shows promising results. The prediction accuracy of the fuzzy model based on a k-fold cross-validation of the data samples shows above 89% performance in predicting health-shock based on the given factors.

The rest of the chapter is organised as follows: Section 6.2 presents the proposed data modelling and analytics approach used for health-shock prediction; Section 6.3 presents the system evaluation; and Section 6.4 provides the summary of this discourse.

6.2 Proposed Data Modelling and Analytics Approach

6.2.1 Data Preprocessing

Here, the variables in the dataset were combined into four main factors (i.e. derived variables) namely: Living Standard (Scott et al., 2005), Health Risk (WHO, 2013), Access (Mattson, 2011), and Income Allocation (GoP, 2013) as shown in Appendix I. Here, it is worth mentioning that we have also used the weightage/scaling approach for data preprocessing as followed in (Alkire and Foster, 2007; Naveed and Islam, 2012). Given below is the description of how these factors were calculated based on information derived from the "HexChange" tool.

6.2.1.1 Living Standard

The identification of distinct dimensions and relevant indicators to measure poverty depends upon how poverty is perceived and effects the personal well-being within a society (Alkire and Foster, 2007; Jamal, 2009; Naveed and Islam, 2012). In this regard, poverty indicators help to identify hidden disparities in health variables such as health status, treatment costs, and state of healthcare delivery across societies with different living standards. Therefore, an accurate approach is needed to measure the living standards considering the poverty indicators (Scott et al., 2005). Here, the optimal selection of measuring approach varies due to their own limitations as well as due to certain factors of the region that effect health variables.

Here, the living standard is derived using the variables in the dataset which include: Nature of Ceiling (NC), Resistance of House against Severe Weather (RW), Disposing-off of food (DF), Disposing-off of garbage (DG), Disposing-off of water (DW) and Water Source (WS). Theses variables were then combined using the following equation:

$$LivingStandard = \frac{NC + RW + DF + DG + DW + WS}{6} \tag{6.1}$$

Here, the nature of the walls of the house was excluded as the majority of the data instances had the same wall type. Furthermore, each of the above mentioned factors were normalized by dividing the value by the maximum value. Finally, an average of these factors was taken.

6.2.1.2 Health Risk

This factor was derived using the variables, which are related to the current health situation of the person. Theses variables include: Minor disease frequency (MD), Severe disease frequency (SD), Dental hygiene level (DH), General hygiene level (GH), and Toilet Facility (TF). The influence of these variables contributing to health risk is not equal, hence in calculating health risk, the variables were given different weights as shown below:

$$HealthRisk = \frac{2 \times MD + 4 \times SD + DH + GH + TF}{9} \tag{6.2}$$

Based on the frequency of the disease, the value was multiplied by the number of the days or weeks/months in the year. The value was then divided by total number of days/weeks/months in year in order to normalize the value of this factor:

$$NC = \begin{cases} \frac{0}{365} & \text{if } NC = 1 \\ \frac{1}{365} & \text{if } NC = 2 \\ \frac{12}{365} & \text{if} NC = 3 \\ \frac{24}{365} & \text{if} NC = 4 \\ \frac{52}{365} & \text{if} NC = 5 \\ \frac{104}{365} & \text{if} NC = 6 \\ \frac{365}{365} & \text{otherwise} \end{cases} \tag{6.3}$$

Similarly, values of DH, GH, and TF were normalized by dividing them with their maximum possible value.

6.2.1.3 Access

This factor referred to the cost of access to the health treatments (Mattson, 2011). It is influenced by the variables; annual cost of travelling health unit (AHU) in case of minor disease and the annual cost of hospital (AH). Mathematically, it can be written as:

$$\begin{aligned} &Access = AHU + AH, \\ &AHU = (CM \times MD) + (MD \times 50), \\ &AH = (CS \times SD) + (SD \times 250). \end{aligned} \tag{6.4}$$

In the above equation, AHU is defined as the sum of travelling cost per visit to the local health unit (CM) for the treatment of minor disease and the doctor fee per visit (which was found to be a fixed approximate cost of 50 rupees per check in that area). Similarly, AH is defined as the sum of the travelling cost per visit to the hospital (CS) for the treatment of the severe and the doctor fee per visit (which was found to be a fixed approximate cost of 250 rupees per check in that area).

Finally, access was normalized by diving it to its maximum value:

$$NormalizedAccess = \frac{Access}{max(Access)} \tag{6.5}$$

6.2.1.4 Income Allocation

This factor referred to the strength the financial situation of the family (GoP, 2013). It is derived from the variables: Annual Income of the family (AI), Owned Land size (LZ), Cost of Food (CF) (which was found to be 10% of family income per person), Cost of House Maintenance (CHM) (which was found to be 20% of the annual income of the family), and Debt (DT). This factor was calculated based on the following equation:

$$IncomeAllocation = AI + LZ - (CF + DT + CHM), \tag{6.6}$$

where

$$AI = \begin{cases} ED \times 15000 \times 12 & \text{if } ED > 0 \\ 5000 \times 12 & \text{otherwise} \end{cases} \tag{6.7}$$

$$CF = \begin{cases} AD \times 0.1 \times ED & \text{if } ED > 0 \\ AD \times 0.1 \times (5000 \times 12) & \text{otherwise} \end{cases} \tag{6.8}$$

$$CHM = \begin{cases} 0.2 \times ED & \text{if } RW < 4 \& ED > 0 \\ 0.2 \times (5000 \times 12) & \text{otherwise} \end{cases} \tag{6.9}$$

Here, AI was calculated by multiplying the number of earning adults (ED) in the family by 15000 rupees, (which was the average monthly income per working person) which was then multiplied by 12 to calculate the annual income. In the case of an adult with no earnings, the minimum income is considered to be 5000 per month which is given to the needy family in the form of zakat and financial aid provided by the government and different non-governmental organizations.

The total annual income is normalized by being divided onto the maximum total annual income in the dataset. This is done in order to normalize the data range for all factors.

6.2.2 Data Labelling

The data was automatically labelled by calculating the estimated risk of health-shock (EHS) as a product of the following equation:

$$EHS = HealthRisk + Access - LivingStandard - IncomeAllocation. \tag{6.10}$$

where Health Risk and Access were both considered to have a positive relationship with health-shock while Living Standard and Income Allocation were considered to have an inverse relationship with health-shock. The values of EHS were normalised by dividing each data instance with the maximum estimated health-shock found in the data. After the data was automatically labelled, it was presented to a field expert to check and make adjustments if needed. The purpose of labelling the data automatically was to facilitate the expert by providing an initial labelling, which they could then amend as needed.

6.2.3 Proposed Fuzzy Linguistic Summarization Approach

Fuzzy Logic Systems (FLSs) provide a transparent and flexible model that allow for the handling of real world information imprecision through the use of linguistic quantifiers such as Poor or High (Mendel, 2001). FLSs represent a methodology for computing with words where linguistic quantifiers described using fuzzy sets are combined with human interpretable If-Then rules (Mendel, 2001). The fuzzy rules convey richer and more easily understandable linguistic summarization (LS) of patterns associating the independent input variables with the dependent target output decisions or states found in the data (Wu et al., 2010). Additionally the extracted fuzzy classification rules rule have quality measures associated with each rule that can be used to measure the strength of patterns found in the data and provide the ability to rank the top rules associated with particular output conditions. We have used a Fuzzy Linguistic Summarization approach (Doctor and Iqbal, 2012; Meunier and Moyse, 2012; Niewiadomski, 2008) consisting of four phases as shown in Fig. 6.1, which is described below:

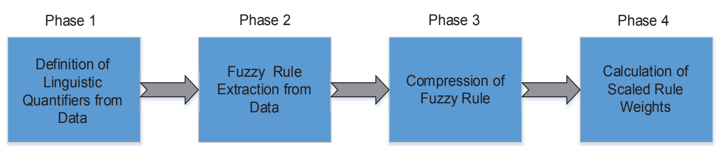


Figure 6.1: Flow diagram showing the phases of the fuzzy LS approach.

6.2.3.1 Definition of Linguistic Quantifiers from Data

In phase 1, the input/output data comprising of the four independent variables and the single dependent variable representing severity of health-shock is mapped to predefined linguistic quantifiers where these derived variables were acquired from the pre-processed data collected during the user study. In numerical and continuous valued data attributes, uncertainties relating to the linguistic quantification over different data values of the attribute suggest the need to use fuzzy sets. This is a generalization of a crisp set that allows the gradual assessment of the membership of an element belonging to a set by using a fuzzy Membership Function (MF) as follows (Zadeh, 1965): Given a domain of discourse $\mathbf{X}$, a fuzzy set $\mathbf{A}$ on $\mathbf{X}$ is a set expressed by a characteristic function $mu_A : \mathbf{X} \rightarrow [0, 1]$ that measures the membership grade of the elements in $\mathbf{X}$ belonging to the set $\mathbf{A}$:

$$\mathbf{A} = \{(x, \mu_A(x)) | \forall\, x \in \mathbf{X}, \mu_A(x) \in [0, 1]\}, \quad (6.11)$$

where $\mu_A(x)$ is called the fuzzy MF of the fuzzy set $\mathbf{A}$.

We divided the preprocessed data into a set of MFs which quantify the values of the data attributes into linguistic labels that partitions the data space into fuzzy regions. Each variable's space is partitioned into five overlapping triangular MFs (Low, Medium to Low, Medium, Medium to High and High) covering the range of the independent and dependent variables, an example of which is shown in Fig. 6.2. This was achieved and further optimised using expert knowledge pertaining to the ranges for each of the variables.

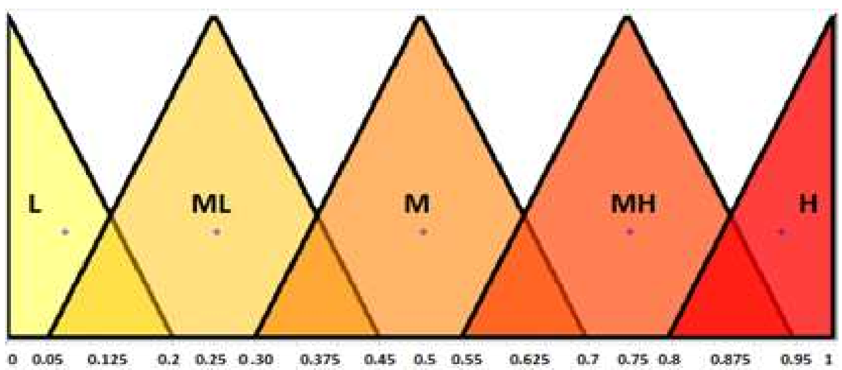


Figure 6.2: Fuzzy Sets for Input and Output Variables.

6.2.3.2 Fuzzy Rule Extraction from Data

In phase 2, fuzzy rule extraction was carried out based on the approach described in the paper by Wang (2003). It is a single-pass method for extracting fuzzy rules from sampled data. The data is mapped to the fuzzy sets for the antecedents and consequents of the rules generated in phase 1. We use the approach to extract multi-input antecedents (for each independent variable) and a single-output consequent (for each dependent/target variable). Combinations of these extracted from the data describe the relationship between y^t and $x^t = (x_1, \cdots, x_n)^t$, that take the following form:

$$\text{IF } x^t \text{ is } \mathbf{A}^q \text{ and } \cdots \text{ and } x_n^t \text{ is } \mathbf{A}_n^q, \text{ THEN } y^t \text{ is } \mathbf{B}^q,$$

where $s = 1, 2, \cdots n$ and n is the number of inputs, $t = 1, 2, \cdots, N$, where N is the number of data instances (Doctor et al., 2005), q is the value of one of the predefined linguistic labels associated to the input or antecedent fuzzy sets $\mathbf{A}$, and the output or consequent fuzzy set $\mathbf{B}$. Using the process described we generate an If-Then proto rule for each data instance. This will result in a profile rulebase comprising of duplicate and contradictory rules.

6.2.3.3 Compression of Fuzzy Rules

In phase 3, the data instance based profile rules are then compressed in order to summarize the data instances into unique end rules. This process involves a modified calculation of two rule quality measures from which we then derive the scaled weight of each unique summarization rule. The quality measures are based on generality measuring the number of data instances supporting each rule (Wu et al., 2010) and reliability measuring the confidence level in the data instances that support each rule (Wu et al., 2010). In our approach the rule generality is measured using fuzzy rule support and the reliability of the rule is based on calculating its confidence.

The fuzzy rule support of a rule is computed as the product of the rule's support and firing strength. The support of a rule refers to coverage of input data instances that map to it (Ishibuchi and Yamamoto, 2005), while it's firing strength measures the degree to which the rule matches those input data instances (Mendel, 2001). The rule's fuzzy support can be used to identify the unique rules with the most frequent occurrences of data instances associated with them, where the data instances most closely map to those rules. The fuzzy support of each rule is scaled based on the total data instances mapping to each output (consequent) set so that the frequencies are scaled in proportion to the number data instances found in each consequent set. The calculation of the scaled fuzzy support for a given uniquely occurring rule is based on the calculation from two appropriate research papers (Ishibuchi and Yamamoto, 2005) and (Doctor and Iqbal, 2012). Following the calculation of the rule support, duplicate instance based proto rules can be identified and eliminated to compress the rule base into a set of M unique and contradictory rules for modelling the data points.

The confidence of a rule measures the rule's validity in describing how tightly data instances are associated with a specific output set. The confidence value range is between 0 and 1. A confidence of 1 implies that the pattern which the rule describes is completely unique to a single output (consequent) set. A confidence of less than 1 implies that the pattern described in the rule occurs in the data associated with more than one output (consequent) set. In this case it should then be interpreted as being best associated with the output set having the highest confidence. The rule scaled confidence is based on the calculation described in the paper by Ishibuchi and Yamamoto (2005) and used in the paper by Doctor and Iqbal (2012).

6.2.3.4 Calculation of Scaled Rule Weights

In phase 4 each rules scaled fuzzy weight is calculated as the product of the scaled fuzzy support and confidence of a rule as shown below:

$$scW_i = FuzzSup \times Conf \tag{6.12}$$

where $FuzzSup$ is the scaled fuzzy support and $Conf$ is the scaled fuzzy confidence. Each of the generated M rules is assigned the scaled fuzzy weight measure scW_i and takes the following form:

$$\text{IF } x^t \text{ is } \mathbf{A}^q \text{ and } \cdots \text{ and } x_n^t \text{ is } \mathbf{A}_n^q\text{, THEN } y^t \text{ is } \mathbf{B}^q,$$

The scaled fuzzy weight measures the quality of each rule in its ability to model the data. It can be used to rank the top rules associated to each output set and choose a single winner rule among compatible rules based on methods for rule weight specification described in the paper by Ishibuchi and Yamamoto (2005) and used in the paper by Doctor and Iqbal (2012).

6.3 System Evaluation

In order to achieve an efficient fuzzy rule based system, two quality aspects should be considered; interpretability and accuracy. Interpretability refers to ability of the model to generate understandable and sensible rules in terms of the real world systems. Accuracy refers to ability of the system to produce a similar response to the real world system (Gacto et al., 2011).

6.3.1 Model interpretability

Interpretability is a subjective property which depends on the expert opinion and could be influenced by different factors such as structure of the fuzzy model, the number of the input variables, the number of the linguistic labels and shape of the fuzzy sets (Gacto et al., 2011). In order to assess the interpretability of the produced fuzzy rules, the rules were reviewed by a team of four experts to give their judgement on whether the rules were understandable and made sense in respect of the health shock risk estimation which resulted from the user study.

Table 6.1 shows a sample of the produced rules where each rule consists of the four antecedents (input variables): living standard, health risk, access and income allocation and one consequent (output variable), which is the heath shock risk estimation. The rules in the table are sorted in descending order by their scaled weight, which expresses its firing strength. The scaled weight was added to increase the interpretability of the rule by providing additional information about how dominant that rule was in terms of representing and modelling the patterns found in the dataset.

Considering the first three rules as examples to explain the interpretability of the generated rules, rule one states that if Living Standard is Medium (M), Health Risk is Medium to Low (ML), Access is Medium (M) and Income Allocation is Medium (M) then Heath Shock is Medium (M) which makes sense as only Health Risk is ML while the three other factors are M making the estimated health-shock risk to be M. Rule two is also understandable suggesting that when Income is MH this decreases the health-shock risk possibly as a result of ML income allocation. Finally rule three, represents the case where health risk is relatively high (MH) and the Access cost to the medical treatment is also high (MH) which is associated with poor Income Allocation (ML) leading to a high (H) Health-Shock risk probability.

Table 6.1: Weighted Fuzzy Rules

	Living Standard	***Health Risk***	***Access Access***	***Income Allocation***		***Heath Shock***	***Scaled Weight***
1.	M	ML	M	M	→	M	0.19463
2.	M	ML	M	MH	→	ML	0.18145
3.	M	MH	MH	ML	→	H	0.12755
4.	M	MH	H	M	→	H	0.09921
5.	MH	ML	M	M	→	ML	0.09677
6.	M	MH	MH	M	→	MH	0.08592
7.	ML	MH	MH	ML	→	H	0.08036
8.	M	MH	M	ML	→	MH	0.07852
9.	M	L	M	M	→	ML	0.0672
10.	M	M	M	M	→	M	0.06486
11.	M	MH	M	M	→	MH	0.05609
12.	M	ML	MH	M	→	M	0.05428
13.	ML	ML	M	M	→	M	0.05399
14.	MH	ML	M	MH	→	ML	0.05161
15.	M	M	MH	L	→	H	0.04762
16.	MH	MH	MH	ML	→	H	0.04762
17.	M	MH	M	L	→	H	0.04762
18.	M	M	MH	M	→	M	0.03874
19.	M	M	MH	ML	→	MH	0.03435
20.	MH	MH	MH	M	→	MH	0.03053
21.	M	MH	MH	MH	→	MH	0.03053

6.3.2 Model classification accuracy

In contrast to interpretability, accuracy is a more objective measure of model performance and there are a number of well defined methods and measures to evaluate the accuracy of the model such as classification and regression which assess the accuracy based on the percentage of the correctly classified data instances in the dataset. The accuracy of the developed fuzzy model was evaluated using the classification measure in two ways; firstly, the ability of the system to accurately classify the health-shock risk on the full dataset (model accuracy) and secondly, the ability of the system. To correctly classify the heath shock risk on unseen data (prediction accuracy) in order to assess that the system performs accurately on a dataset which it was not trained on.

In order to evaluate the modelling accuracy of the fuzzy model on the full dataset, the generated fuzzy rules were applied on the data instances in the dataset and the estimated health-shock was compared with the actual instance labels in order calculate the percentage of correctly classified data instances in the dataset as shown below:

$$acc_j = \frac{1}{h} \sum_{v_i, y_i \in D_k} \sigma(v_i, y_i) \tag{6.13}$$

Where D is the full data set of size h. $\sigma(v, y) = 1$ if $v = y$ and 0 otherwise. v_i is the predicted value of the instance i and y_i is the actual value of the instance i, where $i = 1toh$. Fig. 6.3 shows that the fuzzy rule based system achieved 0.97 modelling accuracy in classifying health-shock risk correctly on the full data set.

In order to evaluate the prediction accuracy of the fuzzy based system on unseen data, k-fold cross-validation was used (Arlot and Celisse, 2009). In k-fold cross-validation, the dataset D is divided into k equal size (of size h items) subsets called folds. The validation process is then carried out for k iterations and in each iteration $j : 1tok$ the subset k_j is held out and called hold-out set D_h. The rest of the subsets are grouped in a training set $D_t = D - k_j$. The accuracy of the model for each fold k_j was calculated as:

$$acc_j = \frac{1}{h} \sum_{v_i, y_i \in D_h} \sigma(v_i, y_i), \tag{6.14}$$

where $\sigma(v, y) = 1$ if $v = y$ and 0 otherwise, v_i is the predicted value of the instance i and y_i is the actual value of the instance i. The final accuracy of the model is calculated by taking the average of the resulting accuracy values for the all iterations as shown below:

$$ACC = \frac{1}{h}\sum_{j=1}^{k} acc_j, \tag{6.15}$$

For the evaluating the fuzzy rule based system a 5-fold cross validation was applied. The dataset D was partitioned into five folds, each representing 20% of the dataset. For each iteration one of the fold subsets was held out and the system was trained on the other folds, representing the remaining 80% of the dataset to extract a set of weighted fuzzy rules. It is standard practice in evaluating machine learning approaches to split the dataset where 80% comprises of the training data and 20% comprise of the hold-out set.

The resulting fuzzy rules of the training process were used to build a fuzzy system to classify each instance in the hold-out sets D_h. The resulting classifications were compared with the actual associated linguistic labels for the data instances in the hold-out sets. This process was repeated five times for the five different folds where for each fold the models classification accuracy was calculated. Fig. 6.3 presents the prediction accuracies for each fold as well as the average prediction accuracy. It can be seen from the table that the highest accuracy of 96% was for the fold three at $k = 3$ and the lowest accuracy of 73% was for the fold two at $k = 2$. The average accuracy of the system was 89%, which shows a relatively good initial prediction accuracy using the proposed fuzzy modelling technique. Given the flexibility of the fuzzy rule based model it is expected that this performance can be improve with more data which can be summarised into more accurate models for predicting health-shock.

6.4 Summary

In this chapter, we have presented a fuzzy rule summarization based predictive model. In the proposed method, data was pre-processed by a team of four experts to derive four measures related to living standards, health risk, accessibility to health facilities and income allocation labelled with the level of health-shock incurred. Furthermore, the pre-processed data was

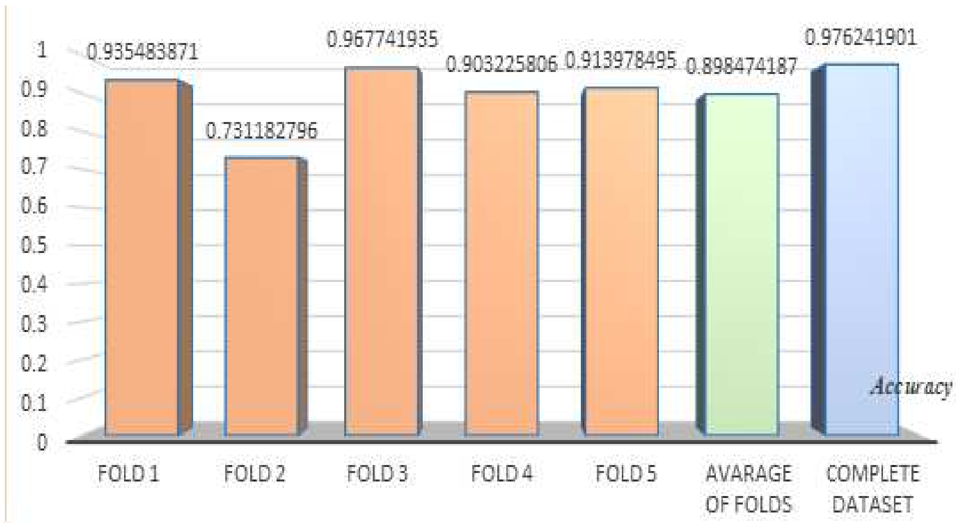


Figure 6.3: Overall model accuracy for prediction health-shock on seen and unseen data.

used to generate a fuzzy rule based classification model for the prediction of health-shock using the fuzzy LS technique which generated an interpretable rule based model to visualize and predict the magnitude of health-shock experienced by individuals. The extracted fuzzy rules used quality measures that determined the strength of each rule in its ability to model the data, which provided stakeholders with a means of ranking and interpreting the quality of the rules. The generated fuzzy model was evaluated based on the interpretability of the rules in their ability to effectively profile the factors affecting different levels of health-shock and the modelling and classification accuracy for predicting health-shock levels from unlabelled data. The results have shown that the generated rules provide sensible and meaningful profiles explaining the factors, corresponding to various levels of health-shock that was also accepted by health experts with knowledge of the health issues affecting the sampled populations. The prediction accuracies of the fuzzy model based on a k-fold cross-validation of the data samples shows that the applied LS approach is also able to achieve good prediction accuracies which can be further improved with larger datasets.

The next chapter will conclude this thesis by presenting the main contributions of this work, its perceived limitations, and outlining the future research directions.

Chapter 7

Conclusions and Future Work

7.1 Introduction

This chapters discuses the contributions made in this thesis by proposing a cloud enabled data analytics and visualization framework for health-shock prediction. Here, the contribution of the thesis is shown by demonstrating how the aims and objectives as set in the first chapters were achieved in order to address the problem of health-shock in general and in the context of Pakistan in particular.

The rest of the chapter is organized as follows: Section 7.2 provides a summary of the research in addition to main contributions of this thesis; Section 7.3 discusses the research limitations; and Section 7.4 presents the future work.

7.2 Research Contributions and Summary

In developing countries, health-shock is more common than other type of shock like economic, political, environmental and law and order. In order to fully comprehend health-shock in developing countries, it is necessary to take socio-economic, geographic, demographic, cultural and climate factors into account. Furthermore, health-shock is one of the variables most associated with poverty worldwide and can yield so many aftershocks such as food insecurity, irreversible malnutrition, termination of schooling, divorce, and missing vaccination, just to name a few. Currently, there is no publicly available dataset that can help to understand, monitor, and predict health-shock based on the historical data. There is an acute shortage of health-shock related data available or a framework to process this information.

Further, the main contributions of this thesis are enumerated as follows:

1. **Development of a context-aware healthcare dataset**: To understand, analyze and predict the possibility of future health-shock, Pakistan's first context-aware healthcare dataset has been developed. For this purpose, a paper based survey questionnaire comprising of 47 features was designed by taking insight from various existing datasets (GoP, 2012, 2015; Naveed and Islam, 2012; Schreiner, 2010). This survey was conducted in the rural areas of Pakistan to collect the data from 1000 households using probability sampling. However, the collected data was redundant and it also proved costly with respect to both time and money.

2. **Development of a cloud enabled mobile application**: In order to address the large human and financial resources associated with data collection at large-scale, a mobile application has been developed. The developed application brings efficiency in health data collection by eliminating data redundancy and errors associated with data entry of the collected survey data. Moreover, it enables the healthcare professionals to create and conduct health surveys on a large-scale with minimal human and financial resources.

3. **Mapping data onto the Iron Triangle**: In order to apply the iron triangle of healthcare (William, 1994) to the socio-economic, geographical, and cultural norms of this region, we have mapped onto it the collected dataset. Such mapping will help in understanding the actual conditions of healthcare in Pakistan.

4. **Development of the data analysis and visualization tool**: To comprehend the hidden patterns in the data, a cloud enabled data analysis and visualization tool has been developed. This tool will help to improve the quality of life for the target population by formulating recommendations for better healthcare policies.

5. **Development of predictive model**: To predict health-shock using qualitative data analysis, a fuzzy rule based summarization technique has been used to develop the one of the first comprehensive health-shock predictive model. Here, it is worth mentioning that the proposed approach is in general in nature, i.e., it can be easily adopted and applied to a wider health informatics dataset.

In general, this research has presented a cloud based data analytics and visualization framework for profiling and predicting health-shock. The framework facilitates the collection of population based socio-economic, geographical, and health related data with control variables using both manual and electronic survey tools which can be easily deployed in remote and rural areas. Large amounts of data can be continuously collected for storage, processing and retrieval on the cloud. The framework was used to carry out a user study comprising of the collection of a unique dataset from 1000 households belonging to 29 villages in rural areas of Pakistan. The data consisted of 47 features, which were pre-processed using health experts to derive four measures related to living standards, health risk, accessibility to health facilities and income allocation labelled with a level of health-shock incurred.

The research further extends in scope by showcasing advance data analytics, visualization, and predictive modelling on this data to save invaluable lives across the world. Knowing the needs and requirements of such a system, a state-of-the-art cloud computing based system has been proposed with Amazon Web Services (AWS), statistical models, and a fuzzy rule summarization based predictive model, to portray a generalized context-aware model utilizing Internet-of-Things (IoT) technology for health-shock modelling and prediction. Here, the biggest achievement was the development of a cloud based infrastructure to capture the first context-aware healthcare dataset based on the socio-economic, cultural and geographic norms of Pakistan. The aim was to then analyze and model such a dataset to understand the relationships between the mentioned factors and their impact on health.

Moreover, empirical assessment and results show that such a model can increase the overall efficiency of modelling and predicting health-shock, reduce the human, and can predict the health-shock for a given population with 89% accuracy, besides several advantages like secure storage of personal health information (PHIs), faster processing, low cost and the use of advanced statistical models, especially, a fuzzy rule summarization based prediction model. Moreover, this research has proposed the first comprehensive approach for health-shock prediction using a fuzzy rule based summarization approach. Here, it is worth mentioning that this work is equally important for general practitioners, governments, standard bodies, United Nations, non-government organizations, and researchers to determine healthcare policies in order to mitigate health-shock.

7.3 Limitations of the proposed work

Apart from the main contributions being highlighted in the previous section, following are some limitations associated with this work:

- Currently, there is no publicly available dataset related to health-shock. Moreover, the proposed analysis and fuzzy based approach rely on a limited dataset. Based on the current dataset, the accuracy of the approach is 89% which can be further improved by collecting more datasets. In this context, the lack of large dataset can be perceived as one of the limitations of the proposed data driven approach.
- Due to the cultural norms of the rural and tribal areas of Pakistan, it is not easy to initiate and/or extend such surveys without the support of government bodies.

7.4 Future Work

The research has demonstrated that large-scale health data analytics will not only help healthcare professionals to create and conduct surveys with minimal human and financial resources but will also help them to understand the socio-economic, environmental and cultural norms that directly or indirectly cause health-shock. This study is one of the first initiatives to analyze and understand the healthcare system and the occurrence of health-shock in rural and tribal areas of Pakistan. In the future, we would like to extend our study to form Pakistan's first publicly available health informatics tool that can be helpful to government and healthcare professionals to form policies and healthcare reforms.

This work can be further developed by deploying it to the masses and collecting thousands of medical surveys using the given survey instrument. A broader survey and analysis will yield a global pattern of health-shock and its economic impact. The work can also be distributed among multiple geographic locations to compare and contrast country wise and/or region wise health-shock conditions. It would be interesting to see its implementation across developed versus developing countries to map the health-shock with other United Nations matrices like the quality of life index and longevity index.

On the technological side, we suggest that this work is extended using the internet of things to connect directly with personal health record systems for a unified health-shock dashboard. We like to extend the work further, developing the health-shock index for different regions and to align with recently formed sustainable development goals of the United Nations, then to release the data free of charge on the internet for other researchers to better the lives of people around the world.

Bibliography

Abraham, A. and Nath, B. (2007). Hybrid Intelligent Systems: A Review of Decade of Research. Technical report, School of Computing and Information Technology, Faculty of Information Technology, Monash University, Australia.

AERA, APA and NCME (2014). *The Standards for Educational and Psychological Testing.*

Agu, E., Pedersen, P., Strong, D., Tulu, B., He, Q., Wang, L., and Li, Y. (2013). The smartphone as a medical device: assessing enablers, benefits and challenges. In *IEEE International Workshop of Internet-of-Things Networking and Control (IoT-NC)*, pages 48–52.

Ahmad, N., Sheikh, M. R., and Saeed, K. (2015). Rural Urban Food Consumption Analysis in Pakistan: Expenditure Elasticities Approach. *Pakistan Journal of Commerce and Social Sciences*, 9(1):159–170.

Ahmed, M. S. and Raja (2010). Tackling cloud security issues and forensics model. In *High-Capacity Optical Networks and Enabling Technologies (HONET)*, pages 190–195.

Ahnn, H. J. and Potkonjak, M. (2013). Toward energy-efficient and distributed mobile health monitoring using parallel offloading. *Journal of medical systems*, 37(5):1–11.

Al-Zaiti, S. S., Shusterman, V., and Carey, M. G. (2013). Novel technical solutions for wireless ECG transmission & analysis in the age of the internet cloud. *Journal of electrocardiology*, 46(6):540–545.

Alabdulatif, A., Khalil, I., and Mai, V. (2013). Protection of electronic health records (EHRs) in cloud. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 4191–4194.

Alagöz, F., Valdez, A. C., Wilkowska, W., Ziefle, M., Dorner, S., and Holzinger, A. (2010). From cloud computing to mobile Internet, from user focus to culture and hedonism: The crucible of mobile health care and wellness applications. In *5th International Conference on Pervasive Computing and Applications (ICPCA)*, pages 38–45.

Alam, K. and Mahal, A. (2014). Economic impacts of health shocks on households in low and middle income countries: a review of the literature. *Globalization and Health*, 10(21):1–18.

Alam, M. (2009). Ageing, Socio-economic Disparities and Health Outcomes: Some Evidence from Rural India. *Indian Journal of Human Development*, 3(1):47–76.

Alcaraz, C., Najera, P., Lopez, J., and Roman, R. (2010). Wireless sensor networks and the internet of things: Do we need a complete integration? In *1st International Workshop on the Security of the Internet of Things (SecIoT10)*.

Alemdar, H. and Ersoy, C. (2010). Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54(15):2688–2710.

Alkire, S. and Foster, J. (2007). Counting and Multidimensional Poverty Measures. Working papers, Oxford Poverty and Human Development Initiative (OPHI), University of Oxford.

Almashaqbeh, G., Hayajneh, T., A. Vasilakos, V., and Bassam, M. J. (2014). QoS-aware health monitoring system using cloud-based WBANs. *Journal of medical systems*, 38(10):1–20.

Altunkaynak, A., zger, M., and akmakci, M. (2005). Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach. *Water Resources Management*, 19(5):641–654.

Amland, R. and Hahn-Cover, K. E. (2014). Clinical Decision Support for Early Recognition of Sepsis. *American Journal of Medical Quality*, page 1062860614557636.

Ammenwerth, E., Brender, J., Nykanen, P., Prokosch, H. U., Rigby, M., and Talmon, J. (2004). Visions and strategies to improve evaluation of health information systems Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International Journal of Medical Informatics*, 73:479–491.

Anderson, G. J. (1997). Clearing the way for physicians' use of clinical information systems. *Communications of the ACM*, 40(8):83–90.

Andreu-Perez, J., Poon, C. C. Y., Merrifield, R. D., Wong, S. T. C., and Yang, G. Z. (2015). Big Data for Health. *IEEE Journal of Biomedical and Health Informatics*, 19(4):1193–1208.

Arlot, S. and Celisse, A. (2009). A survey of cross-validation procedures for model selection.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805.

Azari, A., Janeja, V. P., and Mohseni, A. (2012). Predicting hospital length of stay (PHLOS): A multi-tiered data mining approach. In *12th International Conference on Data Mining Workshops (ICDMW)*, pages 17–24. IEEE.

Balasubramanian, A. (2008). *Health Economics in India*. New Century Publications, New Delhi, 1st edition.

Baldwin, M. L., Zeager, L. A., and Flacco, P. R. (1994). Gender differences in wage losses from impairments. *Journal of Human Resources*, 29(3):865–887.

Barua, M., Liang, X., Lu, R., and Shen, X. (2011). ESPAC: Enabling Security and Patient-centric Access Control for eHealth in cloud computing. *International Journal of Security and Networks*, 6(2-3):67–76.

Basu, S., Karp, A. H., Li, J., Pruyne, J., Rolia, J., Singhal, S., Suermondt, J., and Swaminathan, R. (2012). Fusion: managing healthcare records at cloud scale. *Computer Society*, 45(11):42–49.

Batts, J. W. (2001). Health, Wealth, and Gender: Do Health Shocks of Husbands and Wives Have Different Impacts on Household Wealth? Working papers, Claremont McKenna College, Michigan Retirement Research Center, University of Michigan.

Bazzani, M., Conzon, D., Scalera, A., Spirito, M., Trainito, C. I., et al. (2012). Enabling the IoT paradigm in e-health solutions through the VIRTUS middleware. In *11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 1954–1959. IEEE.

Benharref, A. and Serhani, M. A. (2014). Novel cloud and soa-based framework for e-health monitoring using wireless bio-sensors. *Journal of Biomedical and Health Informatics*, 18(1):46–55.

Bernstein, J. (2009). Impact of the economy on health-care. *Changes in Health Care Financing and Organization (HCFO)*, pages 1–8.

Beyer, M. A. and Laney, D. (2012). The importance of big data: A definition. *CT: Gartner Stamford*.

Bhange, G. S. and Hiray, S. R. (2012). Dental patient records maintenance on cloud computing using data protection model. In *Proceedings of the CUBE International Information Technology Conference*, pages 526–531. ACM.

Bhargava, A., Jamison, D. T., Lan, L., and Murray, C. J. L. (2001). Modelling the Effects of Health On Economic Growth. *Journal of Health Economics*, 20:423–440.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowets, N. A., and Sudman, S. (2004). *Measurement Errors in Surveys*. John Wiley and Sons.

Billings, J., Mijanovich, T., Dixon, J., Curry, N., Wennberg, D., Darin, B., and Steinort, K. (2006). Case finding algorithms for patients at risk of re-hospitalisation. PARR1 and PARR2. *Kings Fund*.

Binnendik, E., Koren, R., and Dror, D. M. (2012). Hardship financing of healthcare among rural poor in Orissa, India. *BMC Health Services Research*, 12:23.

Brill, E. (2003). Processing natural language without natural language processing. In *Computational linguistics and intelligent text processing*, pages 360–369. Springer.

Bryman, A. (1996). *Quantity and Quality in Social Research (Contemporary Social Research)*. Taylor and Francis.

Burgun, A., Bot, G., Fieschi, M., and Le, P. B. (1999). Sharing knowledge in medicine: semantic and ontologic facets of medical concepts. In *International Conference on Systems, Man, and Cybernetics*, volume 6, pages 300–305. IEEE.

Burkhauser, R. V. and Daly, M. C. (1996). *Employment and Economic Well-Being Following the Onset of a Disability: The Role for Public Policy*. Disability, Work, and Cash Benefits.

Burkhauser, R. V. and Mary, C. D. (1994). The economic consequences of disability a comparison of german and american people with disabilities. *Journal of Disability Policy Studies*, 5:25–52.

Burns, L. R. (2014). A system perspective on India's health-care industry. *Global Economy, Health-care reform, India, World Development*.

Callaway, B. and Ghosal, V. (2012). Adoption and Diffusion of Health Information Technology: The Case of Primary Care Clinics. CESifo Working paper: Industrial Organization No. 3925, Leibniz Institute for Economic Research at the University of Munich.

Castillejo, P., Martinez, J., Rodríguez-Molina, J., and Cuerva, A. (2013). Integration of wearable devices in a wireless sensor network for an E-health application. *Wireless Communications*, 20(4):38–49.

Chalmers, A. F. (2004). *What is this thing called science?* Open University Press, 4th edition.

Chard, K., Russell, M., Lussier, Y. A., Mendonça, E. A., and J. C, S. (2011). A cloud-based approach to medical NLP. In *AMIA Annual Symposium Proceedings*, volume 2011, page 207. American Medical Informatics Association.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., and Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine*, 144(10):742–752.

Chen, L. and Hoang, D. B. (2011). Novel data protection model in healthcare cloud. In *13th International Conference on High Performance Computing and Communications (HPCC)*, pages 550–555. IEEE.

Chen, Y., Lu, J., and Jan, J. (2012). A secure EHR system based on hybrid clouds. *Journal of medical systems*, 36(5):3375–3384.

Cheng, C., Stokes, T. H., and Wang, M. D. (2011). caREMOTE: the design of a cancer reporting and monitoring telemedicine system for domestic care. In *Annual International Conference of the Engineering in Medicine and Biology Society, EMBC*, pages 3168–3171. IEEE.

Choi, K., Chung, S., Rhee, H., and Suh, Y. (2010). Classification and Sequential Pattern Analysis for Improving Managerial Efficiency and Providing Better Medical Service in Public Healthcare Centers. *Healthcare Informatics Research*, 16(2):67–76.

Christin, D., Reinhardt, A., Mogre, P. S., and Steinmetz, R. (2009). Wireless sensor networks and the internet of things: Selected challenges. *Proceedings of the 8th GI/ITG KuVS Fachgespräch Drahtlose Sensornetze*, pages 31–34.

Christoph, J., Griebel, L., Leb, I., Engel, I., Köpcke, F., Toddenroth, D., Prokosch, H. U., Laufer, J., Marquardt, K., and Sedlmayr, M. (2014). Secure Secondary Use of Clinical

Data with Cloud-based NLP Services: Towards a Highly Scalable Research Infrastructure. *Methods of Information in Medicine*, 54(3):276–282.

Chung, W. Y., Lee, Y. D., and Jung, S. J. (2008). A wireless sensor network compatible wearable u-healthcare monitoring system using integrated ECG, accelerometer and SpO2. In *International Conference of the Engineering in Medicine and Biology Society, EMBS.*, pages 1529–1532. IEEE.

Combes, C., Kadri, F., and Chaabane, S. (2014). Predicting hospital length of stay using regression models: Application to emergency department. In *10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM.*

Connolly, S., Wooledge, S., and Aster, T. (2013). Harnessing the value of big data analytics.

Courtney, M. (2013). Puzzling out big data [Information Technology Analytics]. *Engineering & Technology*, 7(12):56–60.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* SAGE Publications Ltd., 4th edition.

Cronbach, L. J. (1971). *Test validation.* American Council on Education, 2nd edition.

Darianian, M. and Michael, M. P. (2008). Smart home mobile RFID-based Internet-of-Things systems and services. In *International Conference on Advanced Computer Theory and Engineering, ICACTE*, pages 116–120. IEEE.

Dash, P. K. et al. (2002). Electrocardiogram monitoring. *Indian J. Anaesth*, 46(4):251–260.

Dixon, B. E., Simonaitis, L., Perkins, S. M., Wright, A., and Middleton, B. (2014). Measuring agreement between decision support reminders: The cloud vs. the local expert. *BMC medical informatics and decision making*, 14(1):31.

Dixon, B. E., Simonaitis, L. S., Goldberg, H. S., Paterno, M. D., Schaeffer, M., Hongsermeier, T., Wright, A., and Middleton, B. (2013). A pilot study of distributed knowledge management and clinical decision support in the cloud. *Artificial intelligence in medicine*, 59(1):45–53.

Djam, X. Y., Wajiga, G. M., Kimbi, Y. H., and Blamah, N. V. (2011). A Fuzzy Expert System for the Management of Malaria. *International Journal of Pure and Applied Science and Technology*, 5(2):84–108.

Doctor, F., Hagras, H., and Callaghan, V. (2005). An Intelligent Fuzzy Agent Approach for Realising Ambient Intelligence in Intelligent inhabited environments. *IEEE Transactions on System, Man and Cybernetics*, 35(1):55–65.

Doctor, F. and Iqbal, R. (2012). An Intelligent Framework for Monitoring Student Performance Using Fuzzy Rule-Based Linguistic Summarization. In *FUZZ-IEEE*, pages 1–8.

Doctor, F., Iqbal, R., and Gorgui-Naguib, R. N. (2014). A fuzzy ambient intelligent agents approach for monitoring disease progression of dementia patients. *Journal of Ambient Intelligence and Humanized Computing*, 5(1):147–158.

Dohr, A., Modre, R. O., Drobics, M., Hayn, D., and Schreier, G. (2010). The internet of things for ambient assisted living. In *International Conference on Information Technology: New Generations (ITNG)*, pages 804–809. Ieee.

Doorslaer, E. V. (2006). Effect of payments for health care on poverty estimates in 11 countries in Asia: An analysis of household survey data. *The Lancet*, 368:1357–1364.

Doorslaer, E. V. (2007). Paying out-of-pocket for health care in Asia: Catastrophic and poverty impact. Working papers, eSocialSciences.

Doukas, C. and Maglogiannis, I. (2011). Managing wearable sensor data through cloud computing. In *Third International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 440–445. IEEE.

Doukas, C. and Maglogiannis, I. (2012). Bringing IoT and cloud computing towards pervasive healthcare. In *Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pages 922–926. IEEE.

Doukas, C., Pliakas, T., and Maglogiannis, I. (2010). Mobile healthcare information management utilizing Cloud Computing and Android OS. In *Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 1037–1040. IEEE.

Drew, B. J., Califf, R. M., Funk, M., Kaufman, E. S., Krucoff, M. W., Laks, M. M., Macfarlane, P. W., Sommargren, C., Swiryn, S., Van, H., and George, F. (2004). Practice Standards for Electrocardiographic Monitoring in Hospital Settings An American Heart Association Scientific Statement From the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: Endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. *Circulation*, 110(17):2721–2746.

Dreze, J. and Srinivasan, P. V. (1997). Widowhood and poverty in rural india: some references from household survey data. *Journal of Development Economics*, 54:217–234.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics and Systems*, 3(3):32–57.

Edelsberg, J., Hagiwara, M., Taneja, C., and Oster, G. (2006). Risk of Venous Thromboembolism Among Hospitalized Medically ill Patients. *American Journal of Health-System Pharmacy*, 63(20):S16–S22.

Elbers, C., Gunning, J. W., and Kinsey, B. (2007). Growth and risk: Methodology and micro evidence. 21:1–20.

Eriksson, H., Raciti, M., Basile, M., Cunsolo, A., Fröberg, A., Leifler, O., Ekberg, J., and Timpka, T. (2011). A cloud-based simulation architecture for pandemic influenza simulation. In *AMIA Annual Symposium Proceedings*, volume 2011, page 364. American Medical Informatics Association.

Ermakova, T. and Fabian, B. (2013). Secret sharing for health data in multi-provider clouds. In *15th Conference on Business Informatics (CBI)*, pages 93–100. IEEE.

Fan, Y. J., Yin, Y. H., Da, L. X., Zeng, Y., and Wu, F. (2014). IoT-based smart rehabilitation system. *Transactions on Industrial Informatics*, 10(2):1568–1577.

Fleming, G., Merwe, M. V. D., and McFerren, G. (2007). Fuzzy expert systems and GIS for cholera health risk prediction in southern Africa. *Environmental Modelling and Software*, 22:442–448.

Fong, E. M. and Chung, W. Y. (2013). Mobile cloud-computing-based healthcare service by noncontact ECG monitoring. *Sensors*, 13(12):16451–16473.

Fuchs, V. R. (2010). Health Care Reform: A Collection of Articles on U.S. Health Care Reform. Policy brief, Princeton: Robert Wood Johnson Foundation.

Fujita, H., Uchimura, Y., Waki, K., Omae, K., Takeuchi, I., and Ohe, K. (2012). Development and clinical study of mobile 12-lead electrocardiography based on cloud computing for cardiac emergency. *Studies in health technology and informatics*, 192:1077–1077.

Gachet, D., De, M. B., Aparicio, F., and Padron, V. (2012). Integrating internet of things and cloud computing for health services provisioning: The virtual cloud carer project. In *Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pages 918–921. IEEE.

Gacto, Jos, M., Alcal, R., and Herrera, F. (2011). Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Information Sciences*, 181(20):4340–4360.

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., Beyene, J., Sam, J., and Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*, 293(10):1223–1238.

Gayat, E., Bodin, A., Sportiello, C., Boisson, M., Dreyfus, J. F., Mathieu, E., and Fischler, M. (2011). Performance evaluation of a noninvasive hemoglobin monitoring device. *Annals of Emergency Medicine*, 57(4):330–333.

Ge, L. Y. Y., Li, W., Rao, W., and Shen, W. (2014). A home mobile healthcare system for wheelchair users. In *18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 609–614. IEEE.

Gill, P., Stewart, K., Treasure, E., and Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. *British Dental Journal*, 204:291–295.

Goncalves, F., Macedo, J., Nicolau, M. J., and Santos, A. (2013). Security architecture for mobile e-health applications in medication control. In *21st International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–8. IEEE.

GoP (2010). Government of Khyber Pakhtunkhwa: Khyber Pakhtunkhwa Health Sector Situation Analysis. Technical report.

GoP (2012). Pakistan poverty scorecard: Assessment of measuring impact of ppaf interventions using pakistan poverty scorecard. pages 1–33.

GoP (2013). Household Integrated Economic Survey (HIES). Technical report, Statistics Division Pakistan Bureau of Statistics Islamabad, Pakistan.

GoP (2013). Rapid Assessment Survey of Children's Involvement in Worst Forms of Child Labour in District Haripur, KPK. Technical report.

GoP (2014). Pakistan economy survey 2013−2014. Economy survey, Government of Pakistan.

GoP (2015). Pakistan poverty alleviation fund (PPAF) livelihoods programme: Implementation manual.

Gorp, P. V. and Comuzzi, M. (2014). Lifelong personal health data and application software via virtual machines in the cloud. *Journal of Biomedical and Health Informatics*, 18(1):36–45.

Gronbaek, I. (2008). Architecture for the Internet of Things (IoT): API and interconnect. In *Second International Conference on Sensor Technologies and Applications.* , pages 802–807. IEEE.

Guan, Z. J. (2013). Somatic data blood glucose collection transmission device for Internet of Things.

Gul, O., Al-Qutayri, M., Hieu, V. Q., and Yeun, C. Y. (2012). Framework of a national level electronic health record system. In *International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM).*, pages 60–65. IEEE.

Gupta, I. and Sankar, D. (2001). Health of the elderly in india: A multivariate analysis. *Discussion Paper Series No. 46*.

G.Zhang, Li, C., Zhang, Y., Xing, C., and Yang, J. (2012). SemanMedical: A kind of semantic medical monitoring system model based on the IoT sensors. In *14th International Conference one-Health Networking, Applications and Services (Healthcom)*, pages 238–243. IEEE.

Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *Intelligent Systems*, 24(2):8–12.

Hameed, W., Azmat, S. K., Bilgrami, M., , and Ishaqe, M. (2011). Determining the factors associated with Unmet need for family planning: A cross-sectional survey in 49 districts of Pakistan. *PJPH*, 1:2281–2290.

Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann, 2nd edition.

Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. Massachusetts: MIT Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, 2nd edition.

Heltberg, R. and Lund, N. (1994). Shocks, Coping, and Outcomes for Pakistan's Poor: Health Risks Predominate. *Journal of Development Studies*, 45:889–910.

Hendrick, E., Schooley, B., and Gao, C. (2013). CloudHealth: Developing a reliable cloud platform for healthcare applications. In *Consumer Communications and Networking Conference (CCNC)*, pages 887–891. IEEE.

Herb, H. K., Chandran, A. C., Rudan, I., and Baqui, A. H. (2012). Care seeking for neonatal illness in low and middle income countries: A systematic review. *PLOS Medicine*.

Hersh, W. R. (2002). Medical informatics: improving health care through information. *Jama*, 288(16):1955–1958.

Hiden, H., Woodman, S., Watson, P., and Cala, J. (2013). Developing cloud applications using the e-science central platform. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(1983):20120085.

Horowitz, J. and Savin, N. (2001). Binary Response Models: Logits, Probits and Semiparametrics. *Journal of Economic Perspectives*, 15:43–56.

Houthooft, R., Ruyssinck, J., Herten, J. V. D., Stijven, S., Couckuyt, I., Gadeyne, B., Ongenae, F., Colpaert, K., Decruyenaere, J., Dhaene, T., et al. (2015). Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artificial intelligence in medicine*, 63(3):191–207.

Howlader, M. R. (2013). Analysing the socio-demographic variables impact on health status of Bangladesh. *Social Science Research Network*.

Hsieh, J. C. and Hsu, M. W. (2012). A cloud computing based 12-lead ECG telemedicine service. *BMC medical informatics and decision making*, 12(1):77.

Hu, Y., Lu, F., Khan, I., and Bai, G. (2012). A cloud computing solution for sharing healthcare information. In *ICITST*, pages 465–470.

Huang, J., Sharaf, M., and Huang, C. T. (2012). A hierarchical framework for secure and scalable ehr sharing and access control in multi-cloud. In *41st International Conference on Parallel Processing Workshops (ICPPW)*, pages 279–287. IEEE.

Huang, J. Q., Hooper, P. M., and Marrie, T. J. (2006). Factors associated with length of stay in hospital for suspected community-acquired pneumonia. *Canadian respiratory journal: journal of the Canadian Thoracic Society*, 13(6):317.

Hussain, S., Bang, J. H., Han, M., Ahmed, M. I., Amin, M. B., Lee, S., Nugent, C., McClean, S., Scotney, B., and Parr, G. (2014). Behavior life style analysis for mobile sensory data in cloud computing through MapReduce. *Sensors*, 14(11):22001–22020.

Iles, R. A. (2011). Demand for Primary Health Care in Rural North India: Jointly Modelling Revealed Preference and Qualitative Choice Intentions. *Discussion Papers in Economics.*

Imadali, S., Karanasiou, A., Petrescu, A., Sifniadis, I., Veque, V., and Angelidis, P. (2012). eHealth service support in IPv6 vehicular networks. In *8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 579–585. IEEE.

In, Z. L. (2014). Patient body temperature monitoring system and device based on Internet of Things.

Innocent, P. R. and John, R. I. (2004). Computer Aided Fuzzy Medical Diagnosis. *Information Science*, 162:81–103.

Iqbal, R., Shah, N., James, A., and Duursma, J. (2011). : From work practices to redesign for usability. *Journal of Expert Systems with Applications*, 38:1182–1192.

Ishibuchi, H. and Yamamoto, T. (2005). Rule Weight Specification in Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 13(4):428–435.

Istepanian, R. (2011). The potential of Internet of Things (IOT) for assisted living applications. In *IET Conference Proceedings*. The Institution of Engineering & Technology.

Istepanian, R., Hu, S., Philip, N., and Sungoor, A. (2011). The potential of Internet of m-health Things m-IoT for non-invasive glucose level sensing. In *Annual International Conference of the Engineering in Medicine and Biology Society, EMBC*, pages 5264–5266. IEEE.

Jafar, T., Haaland, H., Rahman, A., Razzak, J., Bilger, M., Naghavi, M., Mokdad, A., and Hyder, A. (2013). Non-communicable diseases and injuries in Pakistan: strategic priorities. *The Lancet*, 381:2281–2290.

Jain, M., Dua, P., Dua, S., and Lukiw, W. J. (2013). Data Adaptive Rule-based Classification System for Alzheimer Classification. *Journal of Computer Science and Systems Biology*, 6(5):291–297.

Jalali, A., A, O. O., and Bell, C. M. (2012). Leveraging Cloud Computing to Address Public Health Disparities: An Analysis of the SPHPS. *Online journal of public health informatics*, 4(3).

Jamal, H. (2009). Estimation of Multidimensional Poverty in Pakistan. Research report no. 79, Social Policy and Development Centre.

Jara, A. J., Zamora-Izquierdo, M. A., and Skarmeta, A. F. (2013). Interconnection framework for mHealth and remote monitoring based on the Internet of Things. *Journal on Selected Areas in Communications*, 31(9):47–65.

Jia, X., Chen, H., and Qi, F. (2012). Technical models and key technologies of E-Health Monitoring. In *14th International Conference on e-Health Networking, Applications and Services (Healthcom).*, pages 23–26. IEEE.

Jian, Z., Zhanli, W., and Zhuang, M. (2012). Temperature measurement system and method based on home gateway.

Kagadis, G. C., Kloukinas, C., Moore, K., Philbin, J., Papadimitroulas, P., Alexakos, C., Nagy, P. G., Visvikis, D., and Hendee, W. R. (2013). Cloud computing in medical imaging. *Medical physics*, 40(7):070901.

Kaletsch, A. and Sunyaev, A. (2011). Privacy engineering: personal health records in cloud computing environments.

Kawaler, E., Cobian, A., Peissig, P., Cross, D., Yale, S., and Craven, M. (2012). Learning to predict post-hospitalization VTE risk from EHR data. In *AMIA Annual Symposium Proceedings*, volume 2012, page 436. American Medical Informatics Association.

Kharat, A. T., Safvi, A., Thind, S. S., and Singh, A. (2012). Cloud Computing for radiologists. *The Indian journal of radiology & imaging*, 22(3):150.

Khattak, H. A., Ruta, M., and Sciascio, E. D. (2014). CoAP-based healthcare sensor networks: A survey. In *11th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pages 499–503. IEEE.

Ko, J., Lu, C., Srivastava, M. B., Stankovic, J., Terzis, A., Welsh, M., et al. (2010). Wireless sensor networks for healthcare. *Proceedings of the IEEE*, 98(11):1947–1960.

Kolici, V., Spaho, E., Matsuo, K., Caballé, S., Barolli, L., and Xhafa, F. (2014). Implementation of a Medical Support System Considering P2P and IoT Technologies. In *Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS).*, pages 101–106. IEEE.

Kondoh, H., Teramoto, K., Kawai, T., Mochida, M., and Nishimura, M. (2012). Development of the regional EPR and PACS sharing system on the infrastructure of cloud computing technology controlled by patient identifier cross reference manager. *Studies in health technology and informatics*, 192:1073–1073.

Koufi, V., Malamateniou, F., and Vassilacopoulos, G. (2010). Ubiquitous access to cloud emergency medical services. In *10th International Conference on Information Technology and Applications in Biomedicine (ITAB)*, pages 1–4. IEEE.

Kuo, A. M. (2011). Opportunities and challenges of cloud computing to improve health care services. *Journal of medical Internet research*, 13(3).

Kupferman, J., Silverman, J., Jara, P., and Browne, J. (2009). Scaling into the cloud. *CS270-advanced operating systems*.

LaMorte, W. W. (2015). An Example of Multiple Linear Regression, Hospital Length of Stay Among Drivers Hospitalized After a Motor Vehicle Collision, Regression Analysis Controlling for Confounding.

Langer, S. G. (2011). Challenges for data storage in medical imaging research. *Journal of digital imaging*, 24(2):203–207.

Laranjo, I., Macedo, J., and Santos, A. (2013). Internet of Things for Medication Control: E-Health Architecture and Service Implementation. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, 2(3):1–15.

Larson, E. C., Goel, M., Boriello, G., Heltshe, S., Rosenfeld, M., and Patel, S. N. (2011). Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 375–384. ACM.

Larson, E. C., Goel, M., Boriello, G., Heltshe, S., Rosenfeld, M., and Patel, S. N. (2012). SpiroSmart: using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 280–289. ACM.

Lazareva, D. (2009). Health Effects of Occupational Change. Working paper no. 129, Centre for Economic and Financial Research at New Economic School.

Leive, A. and Xu, K. (2008). Coping with out-of-pocket health payments: empirical evidence from 15 African countries. *Bulletin of the World Health Organization*, 86:849–856.

Letouzé, E. (2012). Big data for development: challenges & opportunities. *UN Global Pulse*, 47.

Liang, S., Zilong, Y., Hai, S., and Trinidad, M. (2011). Childhood autism language training system and Internet-of-Things-based centralized training center.

Lijun, Z. (2013). Multi-parameter medical acquisition detector based on Internet of Things.

Lin, D. Y. (2013). Integrated Internet of Things application system for prison.

Lindberg, B., Nilsson, C., Zotterman, D., Sderberg, S., and Skr, L. (2013). Using Information and Communication Technology in Home Care for Communication between Patients, Family Members, and Healthcare Professionals: A Systematic Review. *International Journal of Telemedicine and Applications*, pages 1–31.

Liu, C. H., Lin, F. O., Chiang, D. L., Chen, T. L., Chen, C. S., Lin, H. Y., Chung, Y. F., and Chen, T. S. (2013a). Secure PHR access control scheme for healthcare application clouds. In *42nd International Conference on Parallel Processing (ICPP)*, pages 1067–1076. IEEE.

Liu, D. and Tsegai, D. W. (2011). The New Cooperative Medical Scheme (NCMS) and its Implications for Access to Health Care and Medical Expenditure: Evidence from Rural China. *ZEF- Discussion Papers on Development Policy No. 155.*

Liu, J., Li, X., Chen, X., Zhen, Y., and Zeng, L. (2011). Applications of internet of things on smart grid in China. In *13th International Conference on Advanced Communication Technology (ICACT), 2011*, pages 13–17. IEEE.

Liu, J., Tang, Z. H., Zeng, F., Li, Z., and Zhou, L. (2013b). Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. *BMC Medical Informatics and Decision Making*, 13(80):1–7.

Liu, J. and Yang, L. (2011). Application of Internet of Things in the community security management. In *Third International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN).*, pages 314–318. IEEE.

Liu, M. L., Tao, L., and Yan, Z. (2012). Internet of Things-based electrocardiogram monitoring system.

Löhr, H., Sadeghi, A. R., and Winandy, M. (2010). Securing the e-health cloud. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 220–229. ACM.

Lounis, A., Hadjidj, A., Bouabdallah, A., and Challal, Y. (2012). Secure and scalable cloud-based architecture for e-health wireless sensor networks. In *21st international conference on Computer communications and networks (ICCCN)*, pages 1–7. IEEE.

Low, C. and Chen, Y. H. (2012). Criteria for the evaluation of a cloud-based hospital information system outsourcing provider. *Journal of medical systems*, 36(6):3543–3553.

Lu, S., Ranjan, R., and Strazdins, P. (2014). Reporting an experience on design and implementation of e-Health systems on Azure cloud. *Concurrency and Computation: Practice and Experience*.

Luo, H., Ci, H. L., Wu, D., Stergiou, N., and Siu, K. C. (2010). A remote markerless human gait tracking for e-healthcare based on content-aware wireless multimedia communications. *Wireless Communications*, 17(1):44–50.

Mahmud, S., Iqbal, R., and Doctor, F. (2014). An integrated framework for the prediction of health shocks. In *Proceedings of The 2nd International Conference on Applied Information and Communications Technology*, ICAICT.

Mahmud, S., Iqbal, R., and Doctor, F. (2015). Context−Aware HealthCare Dataset - A case study from Pakistan. In *The Fifth International Conference on Advances in Information Mining and Management*, IMMM, pages 77–82.

Mainetti, L., Patrono, L., and Vilei, A. (2011). Evolution of wireless sensor networks towards the internet of things: A survey. In *19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6. IEEE.

Makar, M., Ghassemi, M., Cutler, D. V., and Obermeyer, Z. (2015). Short-Term Mortality Prediction for Elderly Patients Using Medicare Claims Data. *International Journal of Machine Learning and Computing*, 5(3):192.

Manekar, A. and Pradeepini, G. (2015). A Review on Cloud Based Big Data Analytics. *ICSES Journal on Computer Networks and Communication*, 1:6–9.

Maratt, J. D., Srinivasan, R. C., Dahl, W. J., Schilling, P. L., and Urquhart, A. G. (2012). Cloud-based preoperative planning for total hip arthroplasty: a study of accuracy, efficiency, and compliance. *Orthopedics*, 35(8):682–6.

Marshall, D. A., Burgos-Liz, L., IJzerman, M. J., Crown, W., Padula, W. V., Wong, P. K., Pasupathy, K. S., Higashi, M. K., Osgood, N. D., et al. (2015). Selecting a Dynamic Simulation Modeling Method for Health Care Delivery Research Part 2: Report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. *Value in health*, 18(2):147–160.

Marwell, M. R. (2013). *Research Design in Qualitative, Quantitative, and Mixed Methods*. SAGE Publications Ltd., 2nd edition.

Massad, E., Ortega, N. R. S., Barros, L. C., and Struchiner, C. J. (2008). *Fuzzy Logic in Action: Applications in Epidemiology and Beyond*. Springer.

Matthews, P. M., Edison, P., Geraghty, O. C., and Johnson, M. R. (2014). The emerging agenda of stratified medicine in neurology. *Nature Reviews Neurology*, 10(1):15–26.

Mattson, J. (2011). Transportation, Distance, and Health Care Utilization for Older Adults in Rural and Small Urban Areas. *Transportation Research Record*, 2265(1):192–199.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Meitzner, M. and Engineer, P. O. INPATIENT VOLUME FORECASTING PREDICTION MODEL.

Mell, P. and Grance, T. (2009). The NIST definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50.

Mendel, J. (2001). *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, Prentice Hall Inc, 1st edition.

Meunier, B. B. and Moyse, G. (2012). Fuzzy Linguistic Summaries: Where Are We, Where Can We Go? In *IEEE Conference on Computational Intelligence for Financial Engineering and Economics*, pages 317–324. IEEE.

Microsoft (2015). Microsoft HealthVault.

Miori, V. and Russo, D. (2012). Anticipating health hazards through an ontology-based, IoT domotic environment. In *Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pages 745–750. IEEE.

Miras, H., Jiménez, R., Miras, C., and Gomà, C. (2013). CloudMC: a cloud computing application for Monte Carlo simulation. *Physics in medicine and biology*, 58(8):N125.

Morgan, G. and Smircich, L. (1980). The Case for Qualitative Research. *Academy of Management Review*, 5(4):491–500.

Mosa, A. S. M., Yoo, I., and Sheets, L. (2012). A systematic review of healthcare applications for smartphones. *BMC medical informatics and decision making*, 12(1):67.

Murdoch, T. B. and Detsky, S. A. (2013). The inevitable application of big data to health care. *Jama*, 309(13):1351–1352.

Narayan, S., Gagné, M., and Safavi-Naini, R. (2010). Privacy preserving EHR system using attribute-based infrastructure. In *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, pages 47–52. ACM.

Nations, U. (2005). *Household Sample Surveys in Developing and Transition Countries*. Studies in Methods (Ser. F). United Nations Publications, 1st edition.

Naveed, A. and Islam, T. (2012). A New Methodological Framework for Measuring Poverty in Pakistan. Working papers, Sustainable Development Policy Institute.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks: SAGE.

Niewiadomski, A. (2008). A Type-2 Fuzzy Approach to Linguistic Summarization of Data. *IEEE Transactions on Fuzzy Systems*, 16(1):198–212.

Nishtar, S. (2011). *Choked Pipes: Reforming Pakistan's Mixed Health System*. Oxford University Press.

Nishtar, S., Bhutta, Z., Jafar, T., Ghaffar, A., Akhtar, T., Bengali, K., Isa, Q., and Rahim, E. (2003). Health reform in Pakistan: A call to action. *The Lancet*, 381:2291–2297.

Nishtar, S., Boerma, T., Amjad, S., Alam, A., Khalid, F., Haq, I., and Mirza, Y. (2013). Pakistan's health system: Performance and prospects after the 18th constitutional amendment. *The Lancet*, 381:2193–2206.

Nkosi, M. and Mekuria, F. (2010). Cloud computing for enhanced mobile health applications. In *Second International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 629–633. IEEE.

Nussbaum, G. (2006). People with disabilities: assistive homes and environments. In *Computers Helping People with Special Needs*, pages 457–460. Springer.

Onik, A., Fakhrul, M., and Anam, K. (2012). A Secured Cloud based Health Care Data Management System. *International Journal of Computer Applications*, 49(12):0975–8887.

Onukwugha, E. (2015). Big data and its role in health economics.

Pal, S. and Palacios, R. (2011). Understanding poverty among the elderly in india: Implications for social pension policy. *Journal of Development Studies*, 47(7):1017–1037.

Palazuelos, D., Diallo, A. B., Palazuelos, L., Carlile, N., Payne, J. D., and Franke, M. F. (2013). User Perceptions of an mHealth Medicine Dosing Tool for Community Health Workers. *JMIR mHealth and uHealth*, 1(1):2–17.

Pang, Z. (2013). Technologies and Architectures of the Internet-of-Things (IoT) for Health and Well-being.

Pang, Z., Tian, J., and Chen, Q. (2014). Intelligent packaging and intelligent medicine box for medication management towards the Internet-of-Things. In *16th International Conference on Advanced Communication Technology (ICACT)*, pages 352–360. IEEE.

Parsons, D., Robar, J. L., and Sawkey, D. (2014). A Monte Carlo investigation of low-Z target image quality generated in a linear accelerator using Varian's VirtuaLinaca). *Medical physics*, 41(2):021719.

Patra, M. R., Das, R. K., and Padhy, R. P. (2012). CRHIS: cloud based rural healthcare information system. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, pages 402–405. ACM.

Pesta, M., Fichtl, J., Kulda, V., Topolcan, O., and Treska, V. (2013). Monitoring of circulating tumor cells in patients undergoing surgery for hepatic metastases from colorectal cancer. *Anticancer Research*, 33(5):2239–2243.

Piette, J. D., Datwani, H., Gaudioso, S., Foster, S. M., Westphal, J., Perry, W., Rodríguez-Saldaña, J., Mendoza-Avelares, M. O., and Marinec, N. (2012). Hypertension management using mobile technology and home blood pressure monitoring: results of a randomized trial in two low/middle-income countries. *Telemedicine and e-Health*, 18(8):613–620.

Piette, J. D., Mendoza-Avelares, M. O., Ganser, M., Muhima, M., Marinec, N., and Krishnan, S. (2011). A preliminary study of a cloud-computing model for chronic illness self-care support in an underdeveloped country. *American journal of preventive medicine*, 40(6):629–632.

Pirahandeh, M. and Kim, D. H. (2012). Co-designing an intelligent doctors-colleagues-patients social network. In *International Conference on Cloud Computing and Social Networking (ICCCSN)*, pages 1–4. IEEE.

Poole, C. M., Cornelius, I., Trapp, J. V., and Langton, C. M. (2012). Radiotherapy Monte Carlo simulation using cloud computing technology. *Australasian Physical & Engineering Sciences in Medicine*, 35(4):497–502.

Poongodi, M., Manjula, L., Pradeepkumar, S., and Umadevi, M. (2012). Cancer prediction technique using fuzzy logic. *International Journal of Current Research*, 4:106–110.

Prasad, V. M. R. N. R. and Prasad, R. (2011). A cooperative Internet of Things (IoT) for rural healthcare monitoring and control. In *2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*, pages 1–6. IEEE.

Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., and Singer, E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Wiley.

Price, R. C., Pettey, W., Freeman, T., Keahey, K., Leecaster, M., Samore, M., Tobias, J., and Facelli, J. C. (2010). SaTScan on a Cloud: On-Demand Large Scale Spatial Analysis of Epidemics. *Online journal of public health informatics*, 2(1).

Puustjärvi, J. and Puustjärvi, L. (2011). Automating remote monitoring and information therapy: An opportunity to practice telemedicine in developing countries. In *IST-Africa Conference Proceedings*, pages 1–9. IEEE.

Qi, X., Kim, H., Xing, F., Parashar, M., Foran, D. J., and Yang, L. (2012). The analysis of image feature robustness using cometcloud. *Journal of pathology informatics*, 3.

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3):221.

Rajkumar, R. and Iyengar, N. C. S. N. (2013). Dynamic Integration of Mobile JXTA with Cloud Computing for Emergency Rural Public Health Care. *Osong public health and research perspectives*, 4(5):255–264.

Rasid, M. F., Musa, W. M. W., Kadir, N. A. A., Noor, M. A., Touati, F., Mehmood, W., Khriji, L., Al-Busaidi, A., Ben, A. M., et al. (2014). Embedded gateway services for Internet of Things applications in ubiquitous healthcare. In *2nd International Conference on Information and Communication Technology (ICoICT)*, pages 145–148. IEEE.

Ratnam, A. K. and Dominic, P. D. D. (2012). Cloud services-Enhancing the Malaysian healthcare sector. In *International Conference on Computer & Information Science (ICCIS)*, volume 2, pages 604–608. IEEE.

Ratnam, A. K., Dominic, P. D. D., and Ramayah, T. (2014). A structural equation modeling approach for the adoption of cloud computing to enhance the Malaysian healthcare sector. *Journal of medical systems*, 38(8):1–14.

Rea, S., Pathak, J., Savova, G., Oniki, A. T., Westberg, L., Beebe, C. E., Tao, C., Parker, C. G., Haug, P. J., Huff, S. M., et al. (2012). Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. *Journal of biomedical informatics*, 45(4):763–771.

Regola, N. and Chawla, N. V. (2013). Storing and using health data in a virtual private cloud. *Journal of medical Internet research*, 15(3).

Rodrigues, J., de la Torre, I., Fernández, G., and López-Coronado, M. (2013). Analysis of the security and privacy requirements of cloud-based electronic health records systems. *Journal of medical Internet research*, 15(8).

Rodriguez-Martinez, M., Valdivia, H., Rivera, J., Seguel, J., and Greer, M. (2012). MedBook: A cloud-based healthcare billing and record management system. In *5th International Conference on Cloud Computing (CLOUD)*, pages 899–905. IEEE.

Rolim, C. O., Koch, F. L., Westphall, C. B., Jerner, W., Fracalossi, A., and Salvador, G. S. (2010). A cloud computing solution for patient's data collection in health care institutions. In *Second International Conference oneHealth, Telemedicine, and Social Medicine*, pages 95–99. IEEE.

Roman, R., Najera, P., and Lopez, J. (2011). Securing the internet of things. *Computer*, 44(9):51–58.

Rosenthal, A., Mork, P., Li, M. H., Stanford, J., Koester, D., and Reynolds, P. (2010). Cloud computing: a new business paradigm for biomedical information sharing. *Journal of biomedical informatics*, 43(2):342–353.

Rosenzweig, M. R. and Binswanger, H. P. (1993). Wealth, weather risk and the composition and profitability of agricultural investments. *Economic Journal*, 103:56–78.

Rudan, I., Lawn, J., Cousens, S., Rowe, A. K., Boschi-Pinto, C., Tomaskovi, L., Mendoza, W., Lanata, C. F., Roca-Feltrer, A., Carneiro, L., Schellenberg, J. A., Polasek, O., Weber, M., Bryce, J., Morris, S., Black, R. E., and Campbell, H. (2005). Gaps in policy relevant information on burden of disease in children: A systematic review. *The Lancet*, 365:2031–2040.

Ruiz, M. N., García, J. M., and Fernández, B. M. (2009). [Body temperature and its importance as a vital constant]. *Revista de enfermeria (Barcelona, Spain)*, 32(9):44–52.

Sabah, M., Servos, D., and Fiaidhi, J. (2011). Developing a secure distributed OSGI cloud computing infrastructure for sharing health records. In *Autonomous and intelligent systems*, pages 241–252. Springer.

Saldarriaga, A. J., Perez, J. J., Restrepo, J., and Bustamante, J. (2013). A mobile application for ambulatory electrocardiographic monitoring in clinical and domestic environments. In *Pan American Health Care Exchanges (PAHCE)*, pages 1–4. IEEE.

Samuel, O. W., Omisore, M. O., Ojokoh, B. A., and Atajeromavwo, E. J. (2013). Enhanced Cloud based Model for Healthcare Delivery Organizations in Developing Countries. *International Journal of Computer Applications*, 74:34–39.

Saunders, M., Lewis, P., and Thornhill, A. (1998). *The Foundations of Social Research: Meanings and Perspectives in the Research Process*. SAGE Publications Ltd.

Saunders, M., Lewis, P., and Thornhill, A. (2011). *Research Methods for Business Students*. Pearson Book, 5th edition.

Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M., and Oliveira, A. (2011). Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation. *Future Internet Assembly*, 6656:431–446.

Schreiner, M. (2010). A simple poverty scorecard for pakistan. *Journal of Asian and African Studies*, 45(3):326–349.

Schweitzer, E. J. (2012). Reconciliation of the cloud computing model with US federal electronic health record regulations. *Journal of the American Medical Informatics Association*, 19(2):161–165.

Scott, K., Steele, D., and Temesgen, T. (2005). *Chapter XXIII: Living Standards Measurement Study Surveys*. Studies in Methods (Ser. F). United Nations Publications, 1st edition.

Sebestyen, G., Hangan, A., Oniga, S., and Gal, Z. (2014). eHealth solutions in the context of Internet of Things. In *Proceedings of the 2014 International Conference on Automation, Quality and Testing, Robotics (AQTR)*, pages 1–6.

Shahamabadi, M. S., B. M. Ali, B., Varahram, P., and Jara, A. J. (2013). A Network Mobility Solution Based on 6LoWPAN Hospital Wireless Sensor Network (NEMO-HWSN). In *Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pages 433–438. IEEE.

Sharma, M., Bai, Y., Chung, S., and Dai, L. (2012). Using risk in access control for cloud-assisted ehealth. In *14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS)*, pages 1047–1052. IEEE.

Sheikh-Nia, S. (2012). *An Investigation of Standard and Ensemble Based Classification Techniques for the Prediction of Hospitalization Duration*. PhD thesis.

Shen, C. P., Jigjidsuren, C., Dorjgochoo, S., Chen, C. H., Chen, W. H., Hsu, C. K., Wu, J. M., Hsueh, C. W., Lai, M. S., Tan, C. T., et al. (2012). A data-mining framework for transnational healthcare system. *Journal of medical systems*, 36(4):2565–2575.

Shih, F. J., Fan, Y., Chiu, C. M., Shih, F. J., and Wang, S. S. (2012). The Dilemma of "To Be or Not To Be": Developing Electronically e-Health & Cloud Computing Documents for Overseas Transplant Patients from Taiwan Organ Transplant Health Professionals' Perspective. In *Transplantation proceedings*, volume 44, pages 835–838. Elsevier.

Shukla, D. P., Patel, S. B., and Sen, A. K. (2014). A Literature Review in Health Informatics Using Data Mining Techniques. *International Journal of Software and Hardware Research in Engineering*, 2(2):123–129.

Siddiqui, A. Z., Hanan, A., Khan, M. K., and Alghamdi, A. S. (2014). Smart environment as a service: Three factor cloud based user authentication for telecare medical information system. *Journal of medical systems*, 38(1):1–14.

Silvaand, L. A. B., Costa, C., and Oliveira, J. L. (2012). A PACS archive architecture supported on cloud services. *International Journal of Computer Assisted Radiology and Surgery*, 7(3):349–358.

Silvaand, L. A. B., Costa, C., and Oliveira, J. L. (2013). DICOM relay over the cloud. *International Journal of Computer Assisted Radiology and Surgery*, 8(3):323–333.

Smith, J. P. (2007). Consequences and predictors of New Health Events: Analysis in the Economics of Aging. Working paper no. 10063, Chicago: University of Chicago Press.

Sofka, M., Ralovich, K., Zhang, J., Zhou, S. K., Comaniciu, D., et al. (2012). Progressive data transmission for anatomical landmark detection in a cloud. *Methods of information in medicine*, 51(3):268.

Stoean, R., Stoean, C., Sandita, A., Ciobanu, D., and Mesina, C. (2015). Ensemble of Classifiers for Length of Stay Prediction in Colorectal Cancer. In *Advances in Computational Intelligence*, pages 444–457. Springer.

Strobl, C., Malley, J., and Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychological Methods*, 14(4):323–348.

Su, C. J. and Chiang, C. Y. (2013). IAServ: An intelligent home care web services platform in a cloud for aging-in-place. *International journal of environmental research and public health*, 10(11):6106–6130.

Subrahmanya, G. R., Sundararaman, K., and Parthasarathi, J. (2010). Dhatri-A Pervasive Cloud initiative for primary healthcare services. In *14th International Conference on Intelligence in Next Generation Networks*.

Suhrcke, M., McKee, M., Arce, R. S., Tsolova, S., and Mortensen, J. (2006). The Contribution of health to the economy in the European Union. *Public Health*, 120:94–1001.

Sun, J. and Reddy, C. K. (2013). Big data analytics for healthcare. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1525–1525. ACM.

Szolovits, P. (1995). Uncertainty and Decision in Medical Informatics. *Methods of Information in Medicine*, 34:111–121.

Szolovits, P., Patil, R. S., and Schwartz, W. B. (1988). Artificial Intelligence in Medical Diagnosis. *Journal of Internal Medicine*, 108:80–87.

Takeuchi, H., Mayuzumi, Y., Kodama, N., and Sato, K. (2012). Personal healthcare system using cloud computing. *Studies in health technology and informatics*, 192:936–936.

Tan, B. and Tian, O. (2014). Short paper: Using BSN for tele-health application in upper limb rehabilitation. In *IEEE World Forum on Internet of Things (WF-IoT)*, pages 169–170.

Tarouco, M. R. L., Bertholdo, M. L., Granville, Z. L., Arbiza, R. M. L., Carbone, F., Marotta, M., Santanna, D., and Cardoso, J. (2012). Internet of Things in healthcare: Interoperatibility and security issues. In *IEEE International Conference onCommunications (ICC)*, pages 6121–6125. IEEE.

Tong, Y., Sun, J., Chow, S. M. S., and Li, P. (2013). Towards auditable cloud-assisted access of encrypted health data. In *Conference on Communications and Network Security (CNS)*, pages 514–519. IEEE.

Tseng, K. C. and Wu, C. C. (2014). An expert fitness diagnosis system based on elastic cloud computing. *The Scientific World Journal*, 2014.

UNDP (2014). Human Development Report . Technical report.

Vasanth, K. and Sbert, J. (2012). Creating solutions for health through technology innovation. Texas Instruments.

Velthoven, M. H. V., Car, J., Zhang, Y., and Maruic, A. (2013). mHealth series: New ideas for mHealth data collection implementation in low and middle income countries. *PLOS Medicine*, 3.

Vicini, S., Sanna, A., and Bellini, S. (2012). A living lab for internet of things vending machines. In *The impact of virtual, remote, and real logistics labs*, pages 35–43. Springer.

Victor, R., Jeffrey, F., and Harris, E. (1976). Review: Who Shall Live? Health, Economics, and Social choice. *The Bell Journal of Economics*, 7:340–343.

Vida, M. M., Lupse, O. S., Stoicu-Tivadar, L., and Bernad, E. (2012). Flexible solution for interoperable cloud healthcare systems. In *MIE*, pages 280–284.

Wagstaff, A. and Doorslaer, E. V. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. 12:921–34.

Wagstaff, A. and Pradhan, M. (2005). Health insurance impacts on health and nonmedical consumption in a developing country. *Policy Research Working Paper 3563*.

Walsh, P., Cunningham, P., Rothenberg, S. J., O'Doherty, S., Hoey, H., and Healy, R. (2004). An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis. *European Journal of Emergency Medicine*, 11(5):259–264.

Wan, D., Greenway, A., Harris, J. G., and Alter, A. E. (2010). Six questions every health industry executive should ask about cloud computing. *Accenture Institute for Health & Public Service Value*.

Wang, L. X. (2003). The MW method completed: A flexible system approach to data mining. *IEEE Transactions on Fuzzy System*, 11(6):768–782.

Wang, W., Li, J., Wang, L., and Zhao, W. (2011). The internet of things for resident health information service platform research. In *International Conference on Communication Technology and Application (ICCTA 2011)*, pages 631–635. IET.

Wang, X., Gui, Q., Liu, B., Jin, Z., and Chen, Y. (2014). Enabling smart personalized healthcare: a hybrid mobile-cloud approach for ECG telemonitoring. *Journal of Biomedical and Health Informatics*, 18(3):739–745.

Weber, R. H. (2011). Accountability in the Internet of Things. *Computer Law & Security Review*, 27(2):133–138.

Wei, L., Heng, Y., , and Lin, W. Y. (2012). Things based wireless data transmission of blood glucose measuring instruments.

Weitz, M. (1980). *Health Shock*. David and Charles, 1st edition.

WHO (2007). Pakistan's Maternal and Child Health Policy and Strategic Framework (2005-2015). Technical report.

WHO (2008). Disease Early Warning System in Pakistan. Technical report.

WHO (2013). Oral Health: A report by FDI World Health Federation. Technical report.

WHO (2014). UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division). Technical report.

WHO (2015). The Millennium Development Goals Report. Technical report.

William, K. (1994). *Medicines Dilemmas: Infinite need versus finite resources*.

Wooten, R., Klink, R., Sinek, F., Bai, Y., and Sharma, M. (2012). Design and implementation of a secure healthcare social cloud system. In *12th International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, pages 805–810. IEEE.

Wu, D., Mendel, J. M., and Joo, J. (2010). Linguistic Summarization Using If-Then Rules. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, FUZZ-IEEE, pages 1–8.

Wu, R., Ahn, G. J., and Hu, H. (2012a). Secure sharing of electronic health records in clouds. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, pages 711–718. IEEE.

Wu, R., Ahn, G. J., and Hu, H. (2012b). Towards HIPAA-Compliant Healthcare Systems in Cloud Computing. *International Journal of Computational Models and Algorithms in Medicine (IJCMAM)*, 3(2):1–22.

Xiao, Y., Chen, X., Wang, L., Li, W., Liu, B., and Fang, D. (2013). An immune theory based health monitoring and risk evaluation of earthen sites with internet of things. In *International Conference on and IEEE Cyber, Physical and Social Computing Green Computing and Communications (GreenCom)*, pages 378–382. IEEE.

Xu, B., Xu, L. D., Cai, H., Xie, C., Hu, J., and Bu, F. (2014). Ubiquitous data accessing method in IoT-based information system for emergency medical services. *Transactions on Industrial Informatics*, 10(2):1578–1586.

Xu, R., Mei, G., Zhang, G., Gao, P., Pepe, A., and Li, J. (2012). TPM: cloud-based tele-PTSD monitor using multi-dimensional information. *Studies in health technology and informatics*, 184:471–477.

Y. Xiaogang, L. Hongjiang, W. J. and Wentao, T. (2011). Realization of comprehensive detection algorithm of electrocardiogram signal at application layer electrocardiogram monitoring Internet of Thing.

Y. Yue-Hong, F. Wu, F. Y. J. L. J. X. C. and Yi, Z. (2014). Remote medical rehabilitation system in smart city.

Yang, G., Xie, L., Mantysalo, M., Zhou, X., Pang, Z., Xu, L. D., Kao-Walter, S., Chen, Q., and Zheng, L. R. (2014). A Health-IoT Platform Based on the Integration of Intelligent Packaging, Unobtrusive Bio-Sensor, and Intelligent Medicine Box. *Transactions on Industrial Informatics*, 10(4):2180–2191.

Yang, N., Zhao, X., and Zhang, H. (2012). A non-contact health monitoring model based on the internet of things. In *Eighth International Conference on Natural Computation (ICNC)*, pages 506–510. IEEE.

Yao, Q., Han, X., Ma, X. K., Xue, Y. F., Yhen, C. J., and Li, J. S. (2014). Cloud-based hospital information system as a service for grassroots healthcare institutions. *Journal of medical systems*, 38(9):1–7.

Yeh, J. Y., Wu, T. H., and Tsao, C. W. (2011). Using data mining techniques to predict hospitalization of hemodialysis patients. *Decision Support Systems*, 50(2):439–448.

Yoo, S., Kim, S., Kim, T., Baek, R. M., Suh, C. S., Chung, C. Y., and Hwang, H. (2012). Economic analysis of cloud-based desktop virtualization implementation at a hospital. *BMC medical informatics and decision making*, 12(1):119.

Yoshida, H., Wu, Y., Cai, W., and Brett, B. (2012). Scalable, high-performance 3D imaging software platform: system architecture and application to virtual colonoscopy. In *Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 3994–3997. IEEE.

You, L., Liu, C., and Tong, S. (2011). Community medical network (cmn): Architecture and implementation. In *Mobile Congress (GMC)*, pages 1–6. IEEE.

Yu, W., Liu, T., Valdez, R., Gwinn, M., and Khoury, M. J. (2013). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making*, 10(16):1–7.

Yu, Z., Thomborson, C., Wang, C., Wang, J., and Li, R. (2012). A cloud-based watermarking method for health data security. In *International Conference on High Performance Computing and Simulation (HPCS)*, pages 642–647.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zao, J. K., Gan, T. T., You, C. K., Chung, C. E., Wang, Y. T., Méndez, S. J. R., Mullen, T., Yu, C., Kothe, C., Hsiao, C. T., et al. (2014). Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology. *Frontiers in human neuroscience*, 8.

Zhang, Q., Cheng, L., and Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, 1(1):7–18.

Zhang, R. and Liu, L. (2010). Security models and requirements for healthcare application clouds. In *3rd International Conference on Cloud Computing (CLOUD)*, pages 268–275. IEEE.

Zhang, X. M. and Zhang, N. (2011). An open, secure and flexible platform based on internet of things and cloud computing for ambient aiding living and telemedicine. In *International Conference on Computer and Management (CAMAN)*, pages 1–4. IEEE.

Zhao, W., Wang, C., and Y.Nakahira (2011). Medical application on internet of things.

Zhu, Q., Wang, R. R., Chen, Q., Liu, Y., and Qin, W. (2010). Iot gateway: Bridgingwireless sensor networks into internet of things. In *8th International Conference on Embedded and Ubiquitous Computing (EUC)*, pages 347–352. IEEE.

Appendix

A Literature Review

A.1 Cloud Computing

References	Approach/Model Proposed	Technique/Methodology	Results/Performance
Hu et al. (2012)	Implementation of e-Health applications with public cloud	Google app engine was used for cloud services.	Efficient data sharing and management
Benharref and Serhani (2014); Lu et al. (2014)	Implementation of Windows Azure	Windows Azure was used for cloud services.	Efficient data sharing and management
Bhange and Hiray (2012); Sabah et al. (2011); Wooten et al. (2012)	Implementation of Amazon EC2	Amazon EC2 was used for cloud services.	Efficient data sharing and management
Low and Chen (2012)	Provider selection evaluation model	Fuzzy Delphi method (FDM) and fuzzy analytical hierarchy process (FAHP) was used for selection.	Selection and evaluation of specific service providers becomes less complex and more robust
Yoo et al. (2012)	An approach of setting up a private cloud to access the hospital information management system (HIMS) on inside connected devices	A virtualization approach was used with fixed virtual desktops consisting of around 400 virtual machines.	A secure way to access HIMS and with a break-even reach in the 4th year after the initial cost
Kuo (2011)	An approach to store data at central location for mutual collaboration	Oracles Exalogic Elastic Cloud and amazon web services were used to store personal health information	An ease of access to personal health information
Kupferman et al. (2009)	Implementation to ensure uninterrupted ubiquitous access with minimal downtimes	Use of clusters with multiple nodes	Reliability of the health-care systems will enhance

Wan et al. (2010)	A study to bottom line priorities in cloud implementation	Implementation of private clouds for data privacy	Identification of the aspects of cloud computing that are beneficial for the organization
Rodrigues et al. (2013)	A study to identify the security risks associated with storing EHRs on third party cloud	A suggestion to have an external standard organization to perform audit of the cloud providers	An effort to improve data security procedures
Chard et al. (2011)	A prototype to extract, process and manage medical data	Authors proposed a cloud-based SaaS NLP prototype	Aid in performing operation on medical data with HIPAA-compliant data security
Liu et al. (2013a)	Identity based encryption (IBE)	Implementation of IBE to provide access control to patient health records (PHRs)	This mechanism makes the complexity of key management easy
Onik et al. (2012)	Cipher text-policy attribute based encryption (ABE)	Implementation of ABE to make security manager module just like human	Ease of access control
Barua et al. (2011)	Patient-centric ABE is proposed	Implementation of access control ABE focusing on patients	Patient have more control over their data access
Narayan et al. (2010) Huang et al. (2012); Lounis et al. (2012); Tong et al. (2013)	ABE and IBE schemes	A mixed usage of ABE and IBE schemes	Detection of normal and emergency access to data
Chen and Hoang (2011)	Cloud-based privacy-aware role based access	It provides role based authenticated access to corresponding operations on medical data	Assurance of authorized access, controllability and data traceability
Ahmed and Raja (2010)	Computer forensic tools (CFT)	Forensic methods are used to minimize the risk of data breach in e-Health systems	With the help of data loggers and sniffers installed on the cloud machines any volatile information can be captured at much ease
Löhr et al. (2010)	Trusted privacy domains (TVD)	The execution environment for individual applications is within isolated secure domains thus authorized users will only be able to access medical data within corresponding TVD	More robust data sharing and user authentication
Kaletsch and Sunyaev (2011)	Online referral and appointment planer (ORAP)	ORAP was implemented to securely transfer information to healthcare specialists	It lacked patient-centric operations such as collaboration tools, action plans, view/update entries and medical history of patients

Kondoh et al. (2012)	Oshidori-Net2	Virtual servers and virtual routers were used to design the server-based computing healthcare system	Cloud computing implementation details were not explicit
Rajkumar and Iyengar (2013)	Concept of peer to peer network for data exchange b/w connected hospital and ambulances	A cloud connected with community hospital provides data storage to share patient data with the nurse in the ambulance	Have potential to reduce the death rate in emergency care resulting from delayed transportation facility
Koufi et al. (2010)	Concept of platform as a service and software as a service to implement cloud emergency medical services	The emergency medical data was stored on multiple nodes within the cloud	Emergency services are provided using cloud
Fujita et al. (2012)	Cloud cardiology	The system used a cloud server to share ECG report simultaneously inside or outside of the hospital	Ease of use to share ECG reports
Fong and Chung (2013)	Mobile cloud-based healthcare service for noncontact ECG monitoring	Application used a client-server architecture for ECG monitoring within a healthcare unit	Ease of use to share ECG
Wang et al. (2014)	Hybrid cloud computing by wearing health sensors	A conceptual model was presented of a hybrid cloud computing environment for data storage to perform processing tasks on cloud	The model was energy efficient for mobile devices to save battery by processing tasks on cloud
Hiden et al. (2013)	e-Science central	Implementation of cloud as platform as a service	A review on different aspects of data security, enforcing workflows, and service execution with data storage services
Hiden et al. (2013)	MOVEeCloud project	Utilized wearable accelerometers to e-Science cloud to collect patient data	Gained important insights by monitoring patient physical activities
Cheng et al. (2011)	caREMOTE project	Mobile devices were used to exchange patient information between telemedicine systems using cloud services on GAE	Provides a more robust patient data security and ease of exchange
Zao et al. (2014)	Online EEG-BCI system	Mobile devices and wireless headsets were used to predict patient cognitive states according to the real life situations using cloud servers	Remote monitoring and reporting services with prediction mechanism

Hussain et al. (2014)	Implementation of predictive analysis environment for smartphones	Predictions were made using smartphone sensory data with four host machines and Hadoop	Identification of routine and ultimate life style patterns
Almashaqbeh et al. (2014)	Cloud based real time health monitoring system (CHMS)	A cloud to use multi hop sensor networks for real time monitoring and messaging	Increase in the quality of service by effective exchange of messages
Silvaand et al. (2012)	Cloud based picture archiving and communication systems (PACS)	A cloud was implemented to store and share the medical images	Ease of image sharing
Silvaand et al. (2012, 2013)	Digital imaging and communications in medicine (DICOM)	A DICOM compliant bridge is used with DICOM services for data sharing	Efficient provision of imaging services among different institutions
Sofka et al. (2012)	A proposed system to exchange large medical images between PACS	Analysis servers were proposed for image exchange between PACS	Overcame transmission delays
Doukas et al. (2010)	A proposed android client application to receive patient data on a server	Server used virtual machine of amazon to access it from 3G and WLAN in android phone	Data sharing became more easy on the go while travelling
Yoshida et al. (2012)	An implementation of a framework for distributed image processing	implementation used multi-core CPUs in a single machine and transmission of processed data to cloud platform	Efficient data processing
Qi et al. (2012)	CometCloud	Implementation of hybrid cloud-grid distribution framework for image texture analysis	Evaluation of the major methods used in image texture analysis
Takeuchi et al. (2012)	A prototype of a dietary system on cloud platform using a mobile device	Implementation of data mining techniques to extract dietary information from cloud data	Feature for dietitians to add comments on diet plans was missing
Siddiqui et al. (2014)	Tele-care medical information system (TMIS)	A smartphone equipped with authentication possibilities was used to connect to TMIS	Increased provision of healthcare services with authenticated access
Gorp and Comuzzi (2014)	MyPHRMachines	An application software was used for data visualization and remote virtual machines were used for data at user end	No need of verbal communication as the patients were able to share virtual machine session with corresponding care givers
Xu et al. (2012)	Automated cloud enabled stress disorder monitoring screen for PostTraumatic Stress Disorder (PTSD)	Tele-PTSD Monitor (TPM) was accessed via the internet by utilizing the power of amazon elastic compute cloud	Enabled patients to measure and see their stress recovery states

Su and Chiang (2013)	Intelligent Aging-in-place Home care Web Services (IAServ)	An agent base environment and knowledge layer with cloud computing services was implemented	Provision of an electronic platform for elderly people to provide them healthcare facilities while residing at home
Tseng and Wu (2014)	iFit platform	An expert cloud was used to exchange fitness data through web services	iFit was able to give fitness suggestions for users to improve daily routine
Ratnam et al. (2014)	Electronic HIMS	Microsoft windows azure was used as a part of cloud architecture	A cost effective healthcare system
Yao et al. (2014)	Cloud based medical service delivery framework (CMSDF)	A cloud based virtual desktop infrastructure was shared with smaller health units as SaaS	Medical information was readily available to all connected health units
Dixon et al. (2013)	Clinical decision support system (CDS)	An application was used to send patient data to cloud server where analysis was performed on the basis of certain set of rules	Local assessments and remotely produced results were analyzed using this.
Amland and Hahn-Cover (2014)	Add-on development for sepsis outbreak monitoring	The add-on was integrated with EHRs by using cloud based services	Continuous monitoring of outbreaks of sepsis in patients
Parsons et al. (2014)	VirtuaLinac web application	Amazon cloud with VirtuaLinac was used for modelling of radiation treatment components	Simulations of radiation dose were implemented with increased processing power
Doukas and Maglogiannis (2011)	Textile wearable platform	The textile platform contained sensors to collect biosignals and data was exchanged on mobile device using Rest services on google cloud.	Google cloud charts were used for data visualization which give important insights about heartbeat.
Ratnam and Dominic (2012)	Implementation of cloud services to connect various medical institutes	Used medical devices as a part of cloud architecture to exchange medical data	Increased processing capability using cloud services
Pirahandeh and Kim (2012)	A social network for medical professionals	It employed a private cloud which connected all of the entities to concerned medical resources with P2P communication and search features.	Patients were able to search and choose doctors easily
Rolim et al. (2010)	Sensor based data collection framework	The sensors were attached with medical equipment and the collected data was stored in a cloud	Data collection became more easy
Saldarriaga et al. (2013)	A mobile application for Android and iOS for ambulatory ECG monitoring.	To acquire ECG data from patients Kardia board was used, then this data was sent to smartphone	Doctors were able to receive ECG waveform and could perform diagnosis

A.2 Internet of Things

Authors	Approach/Model Proposed	Technique/Methodology	Results/Performance
Luo et al. (2010); Nussbaum (2006)	Home monitoring system for elderly people	Sensors are used to collect data in combination with IoT technology and smartphones	Ultimately helped to reduce hospital readmissions of elderly people
Liu and Yang (2011)	Smart grids	Smart grids with IoT technologies were implemented for data collection and data mining.	Reduced energy consumption and lower infrastructure cost
Darianian and Michael (2008); Liu and Yang (2011)	Concept of smart buildings	Implementation proposed to use sensors and actuators with buildings	Efficient surveillance and monitoring in real time
Roman et al. (2011)	Implementation of privacy and data ownership policies for authentication	Smart devices will require authentication and permissions of the object to interact with	Ensured data security
Weber (2011)	Proposal of stakeholder based shared governance structure	Implementation of compliance policies for each stakeholder accessing medical data	Avoid the dominance of a single group in power regarding data access
Weber (2011)	Concept of global accountability	According to compliance, global accountability and law enforcement check is proposed	Data security and accountability will enhance
Alemdar and Ersoy (2010); Ko et al. (2010)	Wireless sensor network (WSN)	WSN used for data collection	Conventional approach to collect data
Chung et al. (2008)	WSN based healthcare services	An implementation of WSN with healthcare services with IoT is studied	WSN applications for medical data are evaluated
Castillejo et al. (2013)	Combination of WSNs and wearable devices	WSNs and wearable devices are integrated with a study of corresponding outcomes	Innovation in the field of healthcare applications
Sebestyen et al. (2014)	Activity recognition system	Implementation of IoT technology with mobile platform to monitor patient activities	Ease of remote monitoring
Jara et al. (2013)	Combination of Bluetooth low energy (BLE) with WBAN	Combination of both with IoT is employed in wearable devices	Ease of remote sharing and monitoring
Gronbaek (2008); Zhu et al. (2010)	IoT healthcare network (IoThNet)	Various implementation structures were studied using cloud platform	Exchange of medical data in a convenient and secure way

Yang et al. (2014)	Data center platforms as middleware	Data center platforms deployed as middleware between business logic layer smart objects	Layered architecture
Pang (2013)	Concept of developing open platforms	Proposed a standardized approach for designing interfaces, both software and hardware for EHRs	Enhanced collaboration and interoperability
Fan et al. (2014)	Automating design methodology (ADM) platform	IoThNet platform was used	Enhanced collaboration and interoperability
Bazzani et al. (2012)	VIRTUS	XMPP instant messaging protocol was used for secure communication in IoThNet	Secure communication even in limited network remote areas
Rasid et al. (2014)	Multi sensor and multi user communication approach	A gateway was used to read health data from routers on the edge. After reading, the captured data was parsed by applying certain algorithms according to defined format	Enhanced collaboration and interoperability
Xu et al. (2014)	Cloud enabled system to access healthcare data over IoThNet	3 layer architecture was implemented consisting of database layer, resource layer and business logic layer	Distributed healthcare data was managed and organized well
Shahamabadi et al. (2013)	Ambient assisted living (AAL) environment	Implementation included artificial intelligence based on IoT environment with security control and communication mechanisms	AAL serves elderly with human like service quality to deal with any unwanted situation
Zhang and Zhang (2011)	Closed loop healthcare services with smart object KIT protocols	Implementation has a gateway on desktop machine to verify different aspects of the proposed platform using KIT protocol	System analysis was conducted to verify quality of service (QoS), interoperability, privacy and security, and data storage
Goncalves et al. (2013)	AAL based medication control	Analyzed secure services for IoT in AAL medication control	Self-care became easy
Prasad and Prasad (2011)	Rural healthcare monitoring system	Monitoring services were provided by utilizing IoT technologies with connected hospitals	Energy efficient and remote healthcare delivery services
You et al. (2011)	Community medical network	Architecture consisted of wireless body area networks (WBANs) with information service layered platform for exchanging medical data between healthcare entities and service providers	Facility to gain useful medical suggestions by specialists remotely

Burgun et al. (1999)	Semantic medical access (SMA)	An approach to use medical ontologies and semantics in healthcare applications based on IoT	Understanding of ontologies and semantics
G.Zhang et al. (2012)	Semantic medical monitoring system	Sensor data is saved on cloud and healthcare applications use medical rule engines to analyze data in the cloud	Ontologies based data analysis
Istepanian (2011)	Glucose sensing system for diabetes	The proposed system monitors changes in glucose levels in blood and identifies change patterns using sensors	Useful for managing diet and medication plans
Guan (2013)	Blood glucose somatic data collection model	The model included a blood glucose collector device, mobile phone/computer along with a processor. IoT based transmission device was used for transmitting data on network	Useful for managing diet and medication plans
Liu et al. (2012)	ECG monitoring system	The system had a portable acquisition transmitter and a receiving processor with an ability to communicate wirelessly	Detection of abnormality in heart data
Y. Xiaogang and Wentao (2011)	ECG signal detection algorithm	This algorithm was designed to be employed at application layer within an IoT network model	Detection of abnormality in heart data
Dohr et al. (2010)	Blood pressure monitoring system	The presented system consisted of a KIT blood pressure meter and KIT mobile phone having NFC interface installed	Efficient blood pressure monitoring of patients
Puustjärvi and Puustjärvi (2011)	Remote blood pressure controlling	A communication model for healthcare units for remote blood pressure monitoring	Remote blood pressure monitoring of patients
Istepanian (2011)	Monitoring body temperature within mobile IoT (m-IoT) system	Various samples of body temperatures were collected using mobile devices and IoT technology	Useful for emergency situations
Jian et al. (2012)	Temperature controlling home gateway	Implementation described the use of infrared radiations as the communication channel to transmit temperature values of users	Cost effective and energy efficient IoT based temperature controlling
Jara et al. (2013)	Wearable pulse oximeter	Oximeter with Bluetooth connectivity feature was connected to the Monere platform directly	Helps to monitor blood oxygen saturation level

Larson et al. (2012)	Low power oximeter	The proposed oximeter was used over IoT network to send data	Assistance to remote located patients
Larson et al. (2011)	Wearable oximeter based on WSN	WSN was adapted to IoT network for oximetry using a wearable	Helps to monitor blood oxygen saturation level
Tan and Tian (2014)	Remote rehabilitation services	Implemented IoT technologies	Assistance to remote located patients
Fan et al. (2014)	Intelligent rehabilitation system	The system was implemented using ontologies based design	Helps elderly people facing aging problems and challenges
Lin (2013)	Rehabilitation healthcare facilities for prisoners	Rehabilitation system used IoT network for remote assistance	Assistance to prisoners within prisons
Liang et al. (2011)	Childhood autism	A training system was proposed to overcome language barriers using IoT	Help in childhood autism
Y. Yue-Hong and Yi (2014)	Concept of smart city	Using IoT sensors in smart buildings smart cities can be designed to aid everyday life	Collective health rehabilitation systems
Pang et al. (2014)	Intelligent packaging approach	Two prototypes were implemented namely, iMedBox and I2Pack to monitored packaging activity wirelessly	Facilitates to avoid discrepancies in medication management
Laranjo et al. (2013)	Controlling medication by using RFID tags	RFID tags used the communication channel provided by IoT network	Designing a medication control system to aid AAL solutions
Ge et al. (2014)	Smart wheelchair management system	The system comprised of sensors and WBANs and communication was done over IoT network	Enormous advantages for physically impaired persons
Kolici et al. (2014)	Peer to peer communication mode for a medical support system	The system implemented a vibration control mechanism to control wheelchairs over IoT network	Enormous advantages for physically impaired persons

A.3 Big Data Analysis

Authors	Approach/Model Proposed	Technique/Methodology	Results/Performance
Marshall et al. (2015)	Dynamic simulation modelling (DSM)	DSM uses multiple factors regarding patients such as patient preferences, patient outcomes and performance of healthcare systems	DSM helps to know aspects of healthcare delivery provided by big data such as decisions regarding healthcare management
Matthews et al. (2014)	A suggestion to use multivariate analysis to identify subgroups	A combination of different datasets was used on the basis of genetic information, electronic health records, and administrative medical information for predictive analysis	Limitations of dataset size were identified and suggestion were given to perform analysis only at population level where sufficient health data is available
Garg et al. (2005)	Clinical decision support applications were proposed	Various applications are studied in this perspective	Helps to make important health decisions based on physicians suggestions
Anderson (1997)	Alert system	On the basis of lab values alerts were generated about the health status	Better monitoring services
Hersh (2002)	Reminder system for doctors	Reminders are given regarding preventive measures if and when required	Awareness applications in emergencies
Chaudhry et al. (2006)	Warnings and reminders system about compliance	Whenever compliance is compromised warning messages can be issued	Awareness for clinicians to avoid any malpractices
Sun and Reddy (2013)	A study conducted to explore factors of introducing big data analytics in healthcare	Use of google big query and map reduce for large datasets	Details regarding new healthcare analytics tools

B Thematic Map: Union councils of district Haripur

C Thematic map: Health Facilities of district Haripur

D Ethical Approval from Partner Institute

BEGUM MEHMOODA
WELFARE TRUST

The Begum Mehmooda Welfare Trust (BMWT) is a non-profit organization registered under the Trust Act of Pakistan. The Trust operates Begum Mehmooda Memorial Hospital (BMMH), located in Desra, Haripur, which is a 20-bed hospital providing much needed medical services to a low-income, impoverished area of Haripur district.

With a modest load of about 10,000 patients per year, Begum Mehmooda Memorial Hospital (BMMH) not only focuses on the quality of service provision but also on transparency and accountability of operations to its community and Trustees. In order to achieve these objectives, BMMH has invested in a state of the art Hospital Management and Information System (HMIS) to support the comprehensive information and management requirements of the hospital, including patient, clinical, ancillary and financial management. The systematic collection of electronic health information, compliant with HL7 and HIPAA standards, about individual patients and our catchment population, we are able to better support the medical needs of the local demographics.

With the basic data and IT infrastructure in place, BMWT is keen on collaborating with the medical and public health researchers to explore and identify the socio-economic, hygiene, environmental and other contextual factors that influence the health of our community. Our objective is to develop interventions, awareness and preventative measures to improve the overall health and disease management in our community. Due to the sensitive nature of the data and subjects, BMWT takes its ethical responsibilities very seriously and have a very strict ethical review process that ensures:

- Risk assessment undertaken by the Trust for each study
- Informed and written consent of any and all participants prior to collection or retrieval of any data
- Disclosure of patients' rights to refuse participation without repercussions
- The anonymisation, security and confidentiality of the data
- Disclosure about the purpose of data collection/retrieval and what will be done to data during and after the study
- Non-traceability of the confidential data to specific organisations or people
- Review of any compensation to participants (justification, control, rationality and objectivity) involved with the study
- Voluntary participation only

BEGUM MEHMOODA
WELFARE TRUST

With our ethical review process in place, BMWT and BMMH fully endorse **Mr. Shahid Mahmud's** research into critical illnesses and socio-economic factors that lead to health shocks in our community. We further confirm that his research work pertaining to BMMH and our patients is fully compliant with our internal and national ethics requirements.

If you have any questions regarding the above, we can be contacted at the following address:

Hina Fysal,
Chief Executive Officer
Begum Mehmooda Welfare Trust
Email :info@bmwt.asia
URL : www.bmwt.Asia
Address : House 13 G , St 2, Phase 6 , Bahria Town Islamabad . Pakistan

D.1 BMWT Consent Form

Consent Form

The purpose of this particular survey, The Multidimensional Poverty Assessment Tool (MPAT), is to obtain information regarding the health issues concerning your household vis--vis your area (village). The survey aims to gather and assess data regarding the factors, which contribute towards the health-shock—especially, for the families residing in the rural areas of Pakistan. Subsequently, the data will facilitate in understanding the characteristics of the effects of health-shock on the individual, family, social, and various governance levels.

Your valuable input will assist in identifying and addressing any (major and/or minor) difficulties faced by the people (like yourself) living in the rural areas of Pakistan with regard to the health services provided to them in such areas. Your cooperation will help in resolving as to how could the access to basic health facilities—especially, in the rural areas of Pakistan be further improved? Your participation will help in identifying and assessing any shortfalls with regard to the present health system in the country, Pakistan.

This is a general household survey. You are requested to kindly answer each and every question as honestly as possible. However, if you find any section(s)/question(s) inappropriate and/or offensive, you may choose not to answer the respective section(s)/question(s).

Your participation is, of course, voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time. If you decide not to participate in this study or if you withdrawal from participating at any time, you will not be penalized.

The procedure involves filling a survey form. Your personal information will be kept confidential, i.e., all of your personal information will be masked and your identity will be kept anonymous. Furthermore, your responses will only be used for the research purpose. Also, the researcher, Shahid Mahmud, accepts the responsibility of ensuring the complete confidentiality of your data.

Respondent Agreement

BMWT survey team has been explained to me. I consent to participate. I have had a chance for my questions to be answered. I know that I may refuse to participate or withdraw at any time without any penalty. Moreover, my withdrawal won't result in the loss of health care benefits that I am otherwise receiving from BMWT. I authorize BMWT and research scholars to use my responses for the research purpose.

Please select your choice below:

Clicking on the "agree" button below indicates that:

- You have ready the above information.
- You voluntarily agree to participate.
- You are at least 18 years of age.

Agree ☐

Disagree ☐

________________ ________________

Respondent Signature Date

________________ ________________

Surveyor Signature Date

D.2 BMWT Survey Form

Survey Form

Enumerator:	Time __:__ to __:__	Date (YYYY/MM/DD): 20__/__/__
District:	Tehsil:	Union Council:
Village:	Respondent's Age:	HH Code:
Consent:	Head of the HH's age:	Gender= M / F / M&F
Marital status: married / single / divorced / widowed	CNIC No.:	GPS coordinates:

1. How many female and male adults (age 15 and older) live and sleep in your home for more than 9 months every year?

   Female adults ——— Male adults ——— Don't know ———

2. How many adults live and work outside your household for more than 9 months every year?

   Number of Adults ———

3. How long does it take, in minutes, for the school-age children (age 5-14) in your household to go to school (by any means: for example, walking, bicycle, scooter, bus)?[*If children attend more than one school, enumerator to record the average time.*]

   Number of minutes = Don't know ———

4. How long does it take, in KM (s), for the school-age children (age 5-14) in your household to go to school (by any means: for example, walking, bicycle, scooter, bus)?[*If children attend more than one school, enumerator to record the average distance.*]

   Number of KM (s) = Don't know ———

5. In the last 12 months, how often has someone in your household been ill (any non-serious illness)?

   Never ☐ Yearly ☐ Monthly ☐ Bi Monthly ☐
   Weekly ☐ BiWeekly ☐ Daily ☐

6. In the last 12 months, how often has someone in your household been seriously ill (meaning they are so ill that they stay in bed, or lying down, for two or more days at a time)?

   Never ☐ Yearly ☐ Monthly ☐ Bi Monthly ☐
   Weekly ☐ BiWeekly ☐ Daily ☐

7. How much time does it take for members of your household to reach the nearest health center which can diagnose simple illness, or treat simple injuries, and prescribe basic medicines?

   Number of minutes = Don't know ———

8. How much distance for nearest health center which can diagnose simple illness, or treat simple injuries, and prescribe basic medicines?

   Number of KM (s) = Don't know ———

9. How much time does it take for members of your household to reach the nearest health center which can diagnose and treat complicated or serious illnesses or injuries (can perform surgery)?

   Number of minutes = Don't know ———

10. How much distance for nearest health center which can diagnose and treat complicated or serious illnesses or injuries (can perform surgery)?

    Number of KM(s) = Don't know ———

11. Is there any facility for child vaccination in your area?

    Yes ——— No ——— Time———-

12. How many times does the polio health worker came in your area in last 12 months?

    Number of visits: ———-

13. How many mothers and infants died during child birth in the last 12 months?

    Mothers ——— Infants ———

14. Who takes care of mother during birth/delivery?

    Personnel (specify): ———

15. What is the primary construction material being used to construct the housing unit's exterior walls? [*Information to be collected by enumerator while in the household (ask only if unable to determine answer visually)*]

    (a) Stone & mortar ☐
    (b) Metal sheeting ☐
    (c) Reinforced concrete ☐
    (d) Brick ☐
    (e) Logs ☐
    (f) Earth ☐
    (g) Mud or earth bricks ☐
    (h) Mud & straw ☐
    (i) Thin wood ☐
    (j) Bamboo ☐
    (k) Mud or earth bricks ☐
    (l) Thin Plastic ☐
    (m) Reeds ☐
    (n) Thick fabric ☐

(o) Thin fabric ☐

(p) Other, Specify: ———

16. What is the primary roofing material being used for construction of the housing unit's main roof?[*Information to be collected by enumerator while in the household (ask only if unable to determine answer visually)*]

    (a) Stone & mortar ☐

    (b) Tiles or shingles ☐

    (c) Synthetic roofing material ☐

    (d) Metal sheeting ☐

    (e) Reinforced concrete ☐

    (f) Thin wood ☐

    (g) Thick wood ☐

    (h) Bamboo ☐

    (i) Thick plastic ☐

    (j) Thin plastic ☐

    (k) Straw or reeds ☐

17. Can your home out-stand strong winds, severe rain, snow or hail without being significantly damage?

    No ☐ With major damage ☐ With minor damage ☐ Yes it can survive ☐ I don't know ☐

18. What type of toilet facility does your household usually use?

    No ☐ Communal ☐ Open pit ☐ Private ☐

19. What does your household usually do with food waste?

(a) Discard close to a house [within 25 meters] ☐

(b) Discard near a house [25 to 75 meters from the house] ☐

(c) Discard far from a house [75 meters or more] ☐

(d) Feed to livestock ☐

(e) Burn it ☐

(f) Feed to pets or guard dogs ☐

(g) Burn it ☐

(h) Compost it ☐

(i) Use for biogas generation ☐

(j) Sell to vendor ☐

(k) It is collected regularly [organized garbage collection within 75 meters of house] ☐

(l) It is collected regularly [organized garbage collection further than 75 meters from house] ☐

(m) Put down drain [piped sewage network] ☐

(n) Use to water crops grown for livestock fodder ☐

(o) Use to water vegetable garden ☐

(p) Other, specify: ———

20. What does your household usually do with garbage (non-food waste)?

(a) Discard close to a house [within 25 meters] ☐

(b) Discard near a house [25 to 75 meters from the house] ☐

(c) Discard far from a house [75 meters or more] ☐

(d) Feed to livestock ☐

(e) Burn it ☐

(f) Feed to pets or guard dogs ☐

(g) Burn it ☐

(h) Compost it ☐

(i) Use for biogas generation ☐

(j) Sell to vendor ☐

(k) It is collected regularly [organized garbage collection within 75 meters of house] ☐

(l) It is collected regularly [organized garbage collection further than 75 meters from house] ☐

(m) Put down drain [piped sewage network] ☐

(n) Use to water crops grown for livestock food ☐

(o) Use to water vegetable garden ☐

(p) Other, specify: ———

21. What does your household usually do with water waste (for example, from bathing, cleaning, toilet)?

(a) Discard close to a house [within 25 meters] ☐

(b) Discard near a house [25 to 75 meters from the house] ☐

(c) Discard far from a house [75 meters or more] ☐

(d) Feed to livestock ☐

(e) Burn it ☐

(f) Feed to pets or guard dogs ☐

(g) Burn it ☐

(h) Compost it ☐

(i) Use for biogas generation ☐

(j) Sell to vendor ☐

(k) It is collected regularly [organized garbage collection within 75 meters of house] ☐

(l) It is collected regularly [organized garbage collection further than 75 meters from house] ☐

(m) Put down drain [piped sewage network] ☐

(n) Use to water crops grown for livestock food ☐

(o) Use to water vegetable garden ☐

(p) Other, specify: ———

22. How often does your family members clean their teeth in a week?

    Never ☐ Rarely ☐ One or two days a week ☐ Most days of the week ☐ Usually once a day ☐ Usually two or three times a day ☐ Don't know ☐

23. How often does your family members clean their hands before eating a meal?

    Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Always ☐ Don't know ☐

24. What is the main source of the water that your household use for drinking, cooking, bathing and cleaning inside the home?

    (a) During the rainy season ————

    (b) During the dry season ————

    (c) During most of the year ————

    (d) Don't know ☐

1. Spring Water 2. Pond 3. Pipe line 4. Private well 5. Communal well 6. Rain Water 7. Private bore 8. Communal bore [*"Private" means used primarily by the household, but may also be shared with 2-4 other households, and is located within 100 meters of the household. "Communal" means it is shared by 5 or more households.*]

25. On average, how much time (in minutes) does it take to collect water for your household's drinking, cooking, bathing and cleaning needs per day? [*If water is collected from a piped supply in the household record "1" minute*]

    During the rainy season ———- During the dry season ———- During most of the year ———- Don't know ———-

26. How much land does your household have for agriculture (for crops, grass, trees, etc.)?

    Hectares:———- Don't know———-

    [*Enumerator to convert local measurement to hectares*]

27. During the last two years, was your household able to make, or buy, enough compost/manure or artificial fertilizer for each growing season?

    Household does not think they need to use compost/manure or fertilizer ☐ No ☐ Rarely ☐ Sometimes ☐ Often ☐ Always ☐

28. Do you have livestock?

    Goats: ———-

    Animals BuffalowCow: ———-

29. What kind of ownership does your household have for your land?

No Land ☐ Private Ownership ☐

Communal Ownership ☐

30. What were the problems that you had to face in past 12 months and what was their severity and frequency?

1^{st}	Event # =		Likely severity=		Likely Frequency=	
2^{nd}	Event # =		Likely severity=		Likely Frequency=	
3^{rd}	Event # =		Likely severity=		Likely Frequency=	
4^{th}	Event # =		Likely severity=		Likely Frequency=	
5^{th}	Event # =		Likely severity=		Likely Frequency=	

- Drought
- Dry spell
- Flood
- Erratic rainfall
- Acid rain
- Frost
- Hail
- Snow or blizzard
- Earthquake
- Volcanic eruption
- Typhoon or hurricane
- Tornado

- Strong wind
- Dust storm
- High temperatures
- Low temperatures
- Subzero temperatures
- Fire
- Insect attack
- Crop pests
- Lack of fertilizer &or too expensive
- Bad seeds
- Soil problems
- Livestock disease
- Irrigation problems
- Labor shortage
- Theft
- Low market prices for crops livestock
- Poor market access
- Family sickness
- Debt
- Local conflict
- National conflict
- Taxes
- Unemployment
- Lose house

- Personal violence
- Corruption
- Imprisonment
- Other, specify: ———

31. If two or three of the five negative events you just mentioned in above response were to occur in the next 12 months, what are the three main ways your household would likely react (cope)?

    Don't know ☐ Primary strategy ——————- Secondary strategy ——————- Tertiary strategy ——————-

    - Seek off-farm work
    - Children help more than usual with household work
    - Ask friends to help with farm labor or business
    - Ask family to help with farm labor or business
    - Reduce healthcare spending
    - Reduce alcohol consumption
    - Reduce meat consumption
    - Reduce fuel consumption
    - Use savings or sell jewellery
    - Sell livestock
    - Sell stored grain
    - Sell durable goods
    - Plant fewer crops next growing season
    - Postpone payment of debts
    - Borrow money from relatives

- Borrow money from friends
- Send children to work outside the household
- Borrow money from bank or other financial service provider
- Borrow money from cooperative or village fund (community-based source)
- Take children out of school so they can work
- Lease farmland
- Sell farmland
- Sell business
- Beg for money/food
- Sell/leave home (live with relatives in area)
- Sell/leave home (move to another area)
- Rely on group insurance
- Rely on private insurance
- Rely on local government
- Rely on national government
- Rely on aid organizations
- Seek technical assistance
- Work two jobs
- Start a business
- Seek medical treatment
- Other, specify: ———

32. In case of natural disaster, how long would it take to re-build your house? [*Record answer in months (for example, 2 years = 24 months)*]

    Don't know ☐ Months = ———-

33. Did your household experience a period of time longer than two weeks when there was not enough food in the past 12 months? (if "yes" then please specify its frequency)?

Never ☐ Yearly ☐ Monthly ☐ Bi-Monthly ☐

Weekly ☐ Bi-Weekly ☐ Daily ☐

34. Amount of debt owed? Don't know ☐ Amount = ———-

35. To whom is the majority of this debt owed?

(a) Relatives ☐

(b) Friends ☐

(c) Village fund ☐

(d) Village government ☐

(e) Rural credit cooperative ☐

(f) Private money lender ☐

(g) Micro-finance institution ☐

(h) Government bank ☐

(i) Private bank ☐

(j) Joint village & bank fund ☐

(k) Joint development project & bank fund ☐

(l) Other, specify: ———

E Features and Scaling of BMWT Dataset

E.1 Features of BMWT Dataset

Sr. No.	Features
1	Subject ID
2	Contact Number
3	Gender
4	ID Card No.
5	Marital Status
6	Age
7	Tehsil
8	Union Council
9	Village
10	GPS Coordinates
11	Adults living in house for more than 9 months in a year (Female)
12	Adults living in house for more than 9 months in a year (Male)
13	Adults earning
14	Distance to schools in Kilometers
15	Distance to schools in Minutes
16	Frequency of minor disease/year
17	Frequency of severe disease/year
18	Distance to basic health unit in Kilometers
19	Distance to basic health unit in Minutes
20	Distance to Hospital in Kilometers
21	Distance to Hospital in Minutes
22	Mid Wife During Birth
23	Distance to Vaccination Center
24	Polio Drops
25	Fatalities During Birth
26	Nature of Walls of House
27	Nature of Ceiling of House

28	Resistance of House against Severe Weather
29	Toilet Facility
30	Disposing off of food
31	Disposing of garbage
32	Disposing off of water
33	Dental Hygiene
34	General Hygiene
35	Water Source (most of the year)
36	Water Source (in dry weather)
37	Time Duration for collecting water for one day
38	Agricultural Land (in canals)
39	Expenses of Manure for Land
40	Domestic animals - Buffaloes/Cows
41	Domestic animals - Goats
42	Ownership of Land
43	Expected Problems (1st, 2nd and 3rd preference wise)
44	Solutions to expected problems
45	Duration for reconstruction of House in case of destruction (in months)
46	Shortage of food
47	Debts

E.2 Scaling of BMWT Features

Marital Status

Marital Status	Scale
Single	1
Married	2
Divorced	3
Widow	4

Frequency of Disease/year (Major/Minor)

Frequency of disease	Scale
Never	1
Yearly	2
Monthly	3
Bi Monthly	4
Weekly	5
Bi Weekly	6
Daily	7

Disposing off of Food, Water, Garbage

Disposing off of Food, Water, Garbage	Scale
Discard close to a house [within 25 meters]	1
Discard near a house [25 to 75 meters from the house	2
Discard far from a house [75 meters or more]	3
Feed to Livestock	4
Burn it	5
Feed to pets or guard dogs	6
Compost it	7
Use for biogas generation	8

Sell to vendor	9
It is collected regularly [organized garbage collection within 75 meters of house]	10
It is collected regularly [organized garbage collection further than 75 meters from house]	11
Put down drain [piped sewage network]	12
Use to water crops grown for livestock fodder	13
Use to water vegetable garden	14

Expected Problems

Expected Problems	Number
Drought	1
Dry spell	2
Flood	3
Acid rain	4
Excessive rainfall	5
Frost	6
Hail	7
Snow or blizzard	8
Earthquake	9
Volcanic eruption	10
Typhoon/hurricane	11
Tornado	12
Strong wind	13
Dust storm	14
High temperatures	15
Low temperatures	16

Subzero temperatures	**17**
Fire	**18**
Insect attack	**19**
Crop pests	**20**
Lack of fertilizer &/or too expensive	**21**
Bad seeds	**22**
Soil problems	**23**
Livestock disease	**24**
Irrigation problems	**25**
Labor shortage	**26**
Theft	**27**
Low market prices for crops/livestock	**28**
Poor market access	**29**
Family sickness	**30**
Debt	**31**
Local conflict	**32**
National conflict	**33**
Taxes	**34**
Unemployment	**35**
Lose house	**36**
Personal violence	**37**
Corruption	**38**
Imprisonment	**39**

Nature of Walls of House

Nature of material	Number
Stone & mortar	1
Metal sheeting	2
Reinforced concrete	3
Brick	4
Logs	5
Earth	6
Mud or earth bricks	7
Mud & straw	8
Thin wood	9
Bamboo	10
Thick plastic	11
Thin plastic	12
Reeds	13
Thick fabric	14
Thin fabric	15

Nature of Ceiling of House

Nature of material	Number
Stone & mortar	1
Tiles or shingles	2
Synthetic roofing material	3
Metal sheeting	4
Reinforced concrete	5
Thin wood	6
Thick wood	7
Bamboo	8
Thick plastic	9
Thin plastic	10
Straw or reeds	11

Resistance of House against Severe Weather

Answer	**Scale**
no it cant	**1**
with major damage	**2**
with minimal damage	**3**
Yes, it can survive	**4**
I don't know	**5**

Toilet Facility

Toilet facillity	**Scale**
no	**1**
communal	**2**
private	**3**

Dental Hygiene

Dental Hygiene	**scale**
Never	**1**
Rarely	**2**
One or two days a week	**3**
Most days of the week	**4**
Usually once a day	**5**
Usually two or three times a day	**6**
Don't know	**7**

General Hygiene

General Hygiene (Hand wash)	**Scale**
Never	**1**
Rarely	**2**
Sometimes	**3**
Often	**4**
Always	**5**
Don't know	**6**

Source of Water

Source of Water	Number
Spring Water	19
Pond	8
Pipe line	2
Private well	9
communal well	11
Rain Water	20
Private bore	5
Communal bore	7

Expenses for Manure

Expenses for Manure	Scale
Household does not think they need to use compost/manure or fertilizer	1
no	2
rarely	3
sometimes	4
often	5
always	6

Ownership of Land

Ownership of Land	Scale
No Land	0
Private Ownership	1
Communal Ownership	2

Solution to expected problem

Solution to expected problem	Number
Seek off-farm work	1
Children help more than usual with household work	2
Ask friends to help with farm labor or business	3
Ask family to help with farm labor or business	4
Reduce healthcare spending	5
Reduce alcohol consumption	6
Reduce meat consumption	7
.Reduce fuel consumption	8
Use savings or sell jewelry	9
Sell livestock	10
Sell stored grain	11
Sell durable goods	12
Plant fewer crops next growing season	13
Postpone payment of debts	14
Borrow money from relatives	15
Borrow money from friends	16
Send children to work outside the household	17
Borrow money from bank or other financial service provider	18
Borrow money from cooperative or village fund (community-based source)	19
Take children out of school so they can work	20
Lease farmland	21
Sell farmland	22
Sell business	23
Beg for money/food	24
Sell/leave home (live with relatives in area)	25
Sell/leave home (move to another area)	26
Rely on group insurance	27
Rely on private insurance	28
Rely on local government	29
Rely on national government	30
Rely on aid organizations	31
Seek technical assistance	32
Work two jobs	33
Start a business	34
Seek medical treatment	35

Shortage of Food

Shortage of Food	**Frequency**
Never	**1**
Yearly	**2**
Monthly	**3**
Bi Monthly	**4**
Weekly	**5**
Bi Weekly	**6**
Daily	**7**

Debts

Debts	**Range**	**Scale**
no debt	**0**	**1**
a little	**1-20000**	**2**
moderate amount	**20001-50000**	**3**
very large	**50001- onwards**	**4**

F Transportation Expenses from Village to Health facilities and Main Hospital

Table .7: Transportation Expenses

Sr. No.	Village	Nearest governmental health facility	Distance from village to nearest health facility	Distance from health facility to main Hospital	Travelling cost from village to health facility	Travelling cost from health facility to main hospital
1	Akhorra	RHC Jabr	7.2 KM	47 KM	4000	9000
2	Badhaar	RHC Jabri	3 KM	47 KM	2400	7400
3	Bandi Kiala	RHC Jabri	5.2 KM	47 KM	1600	6600
4	Serbaroot	RHC Jabri	16.5 KM	47 KM	8000	13000
5	Shah Kabal	RHC Jabri	31 KM	47 KM	7000	12000
6	Tehal	BHU Nallah	3.4 KM	35 KM	2200	7200
7	Darkote	BHU Nallah	9.9 KM	35 KM	3000	8000
8	Seri	BHU Nallah	17.5 KM	35 KM	1400	6400
9	Bandi Kiala	BHU Nallah	2.5 KM	35 KM	1200	6200
10	Kohala Bala	BHU Barkot	2.4 KM	45 KM	2000	7000
11	Darkote	BHU Barkot	2.6 KM	45 KM	2400	7400
12	Kohmal	BHU Barkot	2 KM	45 KM	3200	8200
13	Ramial	BHU Barkot	2 KM	45 KM	3600	8600
14	Jalees	BHU Barkot	8 KM	45 KM	5000	10000
15	Banth	BHU Barkot	17.5 KM	45 KM	6000	11000
16	Hariala	BHU Barkot	5.5 KM	45 KM	3200	8200
17	Kharian	BHU Barkot	7 KM	45 KM	2400	7400
18	Neelaan	BHU Barkot	5 KM	45 KM	6000	11000
19	Kotla	BHU Barkot	10 KM	45 KM	7000	12000
20	Makhrial	BHU Barkot	7 KM	45 KM	1600	6600
21	Jandi	RHC Halli	7 KM	20 KM	3600	8600
22	Desra	RHC Halli	5 KM	20 KM	1200	8200
23	Beesbaan	RHC Halli	5.5 KM	20 KM	2400	9400
24	Baghpur Dheri	RHC Halli	8 KM	20 KM	4400	11400
25	Khoi Kamman	RHC Halli	22 KM	20 KM	5600	12600
26	Bharrey	BHU Najafpur	5 KM	38 KM	7000	14000
27	Khoi Bagran	BHU Najafpur	3 KM	38 KM	4400	11400
28	Pakshahi	BHU Najafpur	2 KM	38 KM	1600	8600
29	Dhunian	BHU Najafpur	2 KM	38 KM	2000	9000

G Thematic map representing the distance and travelling cost

G.1 From villages to BHU Barkot (roundtrip) and to the main hospital

G.2 From villages to BHU Hali (roundtrip) and to the main hospital

G.3 From villages to BHU Najfpur (roundtrip) and to the main hospital

G.4 From villages to BHU Nullah (roundtrip) and to the main hospital

G.5 From villages to RHC Jabri (roundtrip) and to the main hospital

H Doctor consulting Fee (PKR) in various cities of Pakistan in 2013–14 (GoP, 2014)

Year	Islam-abad	Karachi Karachi	Lahore Lahore	Faisal-labad	Rawal-pindi	Sukkur Sukkur	Gujra-nwala	Hyder-abad	Pesh-awar	Quetta Quetta	Avg.
1973	15	15	10	10	15	10	10	10	20	10	12.5
1974	18.75	20	15	15	20	16	15	20	20	17.5	17.73
1975	20	25	15	20	22.5	17.5	15	20	20	25	20
1976	23.75	27.75	17.5	20	27.19	20	20	23.75	23.13	28.13	23.12
1977	35	25	20	20	35	20	20	28.75	25	35	26.38
1978	22.5	34	20	20	35	20	20	32.14	33.13	40	27.68
1979	-	48	28.33	40	45	35	20	33.75	35	35	32.01
1980	50	54.44	47.5	40	50	35	32	35	37.5	37.5	41.89
1981	50	60	47.5	70	50	25	32	36	50	32.5	45.3
1982	60	60	50	30	50	40	32	50	12	37.5	42.15
1983	60	60	50	50	50	50	-	58.75	12	37.5	42.83
1984	55	36.11	10	20	50	50	20	45	15.63	45	34.67
1985	50	30	10	20	50	35	32	55	20	45	34.7
1986	50	26.39	14.17	20	50	30	32	55	20	45	34.26
1987	26.25	26.7	24.29	20	25.42	30	32	55	20	46.25	30.59
1988	26.25	26.54	24.29	20	25.42	30	32	50	20	67	32.15
1989	26.88	25.91	24.29	48.33	25.42	30	32	50	20	67	34.98
1990	26.88	26.54	30	51.67	25.83	35	32.5	50	22.5	57	35.79
1991	27.5	27.09	24.64	42	26.67	40	32.5	50	22.5	60	35.29
1992	27.5	26.49	24.64	31.67	29.17	75	32.5	66.67	22.5	52.5	38.86
1993	27.5	28.85	27.14	32.54	29.17	75	43.75	80	27.5	52.5	42.4
1994	27.5	31	24.64	32.5	29.17	70	40	65	30	82.5	43.23
1995	27.5	32.24	30	37.5	30	75	40	65.71	30	90	45.79
1996	32.5	31.88	27.86	30	30	55	40	53	30	80	41.02
1997	32.5	31.88	27.86	35	30.83	60	40	46.25	30	80	41.43
1998	33.44	31.6	33.21	35	30	30	40	33.75	30	107.5	40.45
1999	33.44	32.17	33.93	35	31.25	30	40	33.75	30	107.5	40.75
2000	33.13	32.4	38.93	40	32.92	30	40	33.75	30	107.5	41.86
2001	33.13	33	41.96	40	33.75	30	40	33.75	43.33	107.5	43.64
2002	33.13	35	41.25	40	33.96	30	50	30	43.33	95	43.17
2003	45	36.35	41.96	40	38.75	30	50	31.25	50	100	46.33
2004	45	36.25	41.96	41.25	38.75	30	50	33	50	100	46.62
2005	46.25	38.08	44.29	41.25	42.08	30	50	33.75	50	100	47.57
2006	55	41.73	52.68	41.25	43.75	50	50	33.75	50	100	51.81
2007	55	55	52.68	43.75	43.75	75	50	50	50	120	59.52
2008	75	80	63.21	75	61.67	75	65	50	100	130	77.49
2009	75	93.85	68.93	75	61.67	75	65	50	100	120	78.45
2010	90	93.85	68.93	75	71.67	100	75	60	125	130	88.95
2011	100	93.85	70	80	85	100	75	68.75	166.67	180	101.93
2012	200	100	70.36	90	110	100	75	80	191.61	200	139.18
2013	146.25	100	100	90	135	100	75	100	225	200	155.59

I Integrating Features of BMWT Dataset

Table .9: Derived Variables. ■ represents the Access, ■ represents the health risk, ■ represents the living standard, and ■ presents the income allocation.

Sr. No.	Features
1	Subject ID
2	Contact Number
3	Gender
4	ID Card No.
5	Marital Status
6	Age
7	Tehsil
8	Union Council
9	Village
10	GPS Coordinates
11	Adults living in house for more than 9 months in a year (Female)
12	Adults living in house for more than 9 months in a year (Male)
13	Adults earning
14	Distance to schools in Kilometers
15	Distance to schools in Minutes
16	Frequency of minor disease/year
17	Frequency of severe disease/year
18	Distance to basic health unit in Kilometers
19	Distance to basic health unit in Minutes
20	Distance to Hospital in Kilometers
21	Distance to Hospital in Minutes
22	Mid Wife During Birth
23	Distance to Vaccination Center
24	Polio Drops
25	Fatalities During Birth
26	Nature of Walls of House

27	Nature of Ceiling of House
28	Resistance of House against Severe Weather
29	Toilet Facility
30	Disposing off of food
31	Disposing of garbage
32	Disposing off of water
33	Dental Hygiene
34	General Hygiene
35	Water Source (most of the year)
36	Water Source (in dry weather)
37	Time Duration for collecting water for one day
38	Agricultural Land (in canals)
39	Expenses of Manure for Land
40	Domestic animals - Buffaloes/Cows
41	Domestic animals - Goats
42	Ownership of Land
43	Expected Problems (1st, 2nd and 3rd preference wise)
44	Solutions to expected problems
45	Duration for reconstruction of House in case of destruction (in months)
46	Shortage of food
47	Debts