Implementation of a breast cancer screening prediction algorithm: a knowledge management approach

A research project funded by the NHS Cancer Screening Programmes

Research Team

Dr Rajeev K Bali
(Principal Investigator)
Dr Vikraman Baskaran
Dr Hisbel Arochena
Prof Raouf Naguib

Knowledge Management for Healthcare (KARMAH) research subgroup,
Biomedical Computing and Engineering Technologies (BIOCORE)
Applied Research Group,
COVENTRY UNIVERSITY,
Health Design and Technology Institute (HDTI),
Coventry University Technology Park, Puma Way,
COVENTRY, West Midlands  CV1 2TT

www.coventry.ac.uk/biocore/
www.coventry.ac.uk/karmah/

September 2009
Acknowledgements

The BIOCORE research team would like to thank Julietta Patnick CBE, Director, NHS Cancer Screening Programmes for funding this research project. Additional thanks are due to Dr Matthew Wallis (now with the Cambridge Breast Unit and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre) and Margot Wheaton (Warwickshire, Solihull & Coventry Breast Screening Unit) for their support.

Disclaimer

This report is the edited output of a PhD research project carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group at Coventry University. It is provided for information only to interested practitioners and for academic research purposes. For further information, please contact Dr RK Bali (r.bali@ieee.org).
Abstract

The Warwickshire, Solihull and Coventry Breast Screening Unit (‘Unit’) is one of the largest screening services in the UK. In collaboration with the Unit, an Artificial Intelligence Attendance (AI-ATT) algorithm was recently developed within Biomedical Computing Research Group (BIOCORE), for the prediction of attendance (starting from the second episode) to the Breast Screening Programme (BSP). This research aims to integrate a new algorithm (entitled “Java-based Attendance prediction by Artificial Intelligence for Breast Screening (JAABS)” within this Unit in order to streamline and automate the invitation process and thus increase screening attendance. The work deploys a hybrid research methodology which incorporates both the qualitative and quantitative paradigms. Efforts had focussed on the formulation of a BSP messaging protocol. The protocol processes the uptake dataset from the second episode upwards and uses it as the input for the prediction process. The non-attendee list produced by the AI-ATT algorithm is segregated into their respective GP groups. Using the proposed framework, a message transmission process will be initiated to send the information to the GPs’ network. Following proper authentication and acceptance, the information is updated on the GP’s database. This can instigate the deployment of measures to improve attendance and potentially result in a further reduction in morbidity and mortality rates due to breast cancer. Knowledge Management (KM) tools and techniques were used to effectively manage the transfer of knowledge between the Unit, GPs and eligible women within a screening episode. In concurrence with the UK government’s “e-Government Interoperability Framework” (e-GIF) policy, and in projecting the importance of the Electronic Data Interchange (EDI) within the BSP structure, this project addresses the limitations of the existing communication path between the BSP and primary care providers. The deliverables of the project include an investigation and evaluation of the suitability of HL7 and XML standards for the proposed messaging protocol, a comparison of the proposed standards with EDIFACT-based protocols to highlight salient features. As part of the research, questionnaires were designed and distributed to evaluate the GPs’ role in reducing the BSP non-attendance. A comprehensive triangulation-based validation of the two algorithms (AI-ATT and JAABS) was also featured as part of the research. The JAABS algorithm proposed includes a new attribute (associated to the distance travelled by the women for a mammogram) as a predicting variable to increase the accuracy of prediction. Based on these evaluations and validations, technical reports were produced and communicated to all stakeholders.
Contents

List of Figures

List of Tables

Introduction 1
Setting the scene 1
Breast Screening Programme in the United Kingdom 3
Challenges being addressed 6
Solution proposed 7
Research Question 8
Research Objectives 9
Overview 12

Proposed Knowledge Exchange 14
Information and knowledge 15
Knowledge creation 15
Knowledge Management (KM) 16
Beyond KM 20
Total Knowledge Management in healthcare (TKMh) 23
Knowledge creation and application of TKMh model for this project 24
Knowledge sharing for this project 28
Proposed twin approach 29
Prediction and post-screening results 30
HL7 Message design 32
Modelling methodology 33
Information analysis 33
Information modelling 35
HL7 v3 Message implementation 36
Challenges faced during message design and implementation phase 37
Proposed protocol 39
Proposed Breast Screening Attendance Messaging Protocol (BSAMP) 39
Conclusion 42

Algorithm for Screening Attendance Prediction 43
Predecessor algorithm employed in exploratory research 43
Artificial Intelligence based Attendance Prediction algorithm (AI-ATT) 44
Analysis of AI-ATT 44
AI-ATT working environment 45
Data pre-processing 48
Data Pre-processing design 48
Module design 49
Variables identified for data pre-processing 50
Data pre-processing module implementation 52
Proposed algorithm 53
Java based Attendance prediction by Artificial Intelligence for
Breast Screening (JAABS) algorithm 53
JAABS design 53
JAABS classes 55
Testing JAABS algorithm with simulated dataset 56
Comparison of AI-ATT with JAABS implementation 59
JAABS evaluation 60
Dataset 60
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

Attribute selection 65
JAABS vs AI-ATT attributes 66
Screening distance as a new predictor attribute 67
JAABS ROC tests 68
RBFN test 68
ANNP test 71
JAABS test 73
Summary 75

Knowledge Sharing: GP’s Role 76
GP intervention 76
Types of intervention 76
Questionnaire-based survey 77
Questionnaire analysis 79
Screening women 80
Prediction 81
Electronic Data Interchange (EDI) 81
GP intervention 82
Additional information on interventions 82
Questionnaire results 82
Survey summary 87

Conclusion and Future Work 88
Conceptual development 89
Concept development - Knowledge management 89
Concept development - Artificial Intelligence (AI) 89
Concept development - Health Informatics 90
Original contribution 90
Objectives accomplished 92
Project deliverables accomplishment list 93
Adopting the proposed protocol 93
Challenges 94
Future work 96
Concluding observations 97

References 100

Appendix: Research Papers Published 113
List of Figures

FIGURE 1 - BREAST ANATOMY © (NATIONAL CANCER INSTITUTE 2005) 1
FIGURE 2 - MAMMOGRAPHY MACHINE AND TWO VIEWS OF MAMMOGRAM 3
FIGURE 3 - BREAST SCREENING UPTAKE FROM 1995-2005 4
FIGURE 4 - CANCER DETECTION IN UK FROM 1995 TO 2005 4
FIGURE 5 - CANCER DEATHS REPORTED IN UK FOR 1995 – 2004 5
FIGURE 6 - TYPICAL WEB SERVICES IMPLEMENTATION FOR THE PROPOSED PROTOTYPE 8
FIGURE 7 - CHAPTERS’ RELATIONSHIP WITH THE RESEARCH CORE AREAS 12
FIGURE 8 - KNOWLEDGE CREATION CYCLE (WITH 1ST AND 2ND GENERATION INFORMATION) 14
FIGURE 20 - KNOWLEDGE CREATION SPIRAL 16
FIGURE 21 - MULTIPLE FACETS OF KNOWLEDGE MODEL IN PROJECT-BASED ORGANISATION 18
FIGURE 22 - ORGANISATIONAL KNOWLEDGE SHARING (INTER AND INTRA MODES) 22
FIGURE 23 - TKMH FRAMEWORK 23
FIGURE 24 - KNOWLEDGE CREATION IN PREDICTING SCREENING ATTENDANCE 24
FIGURE 25 - TKMH FRAME WORK APPLIED FOR CURRENT RESEARCH 25
FIGURE 26 - SCREENING ATTENDANCE PREDICTION THROUGH ANN 27
FIGURE 27 - KM-BASED ARCHITECTURE FOR INCREASING SCREENING ATTENDANCE 28
FIGURE 28 - INCREASING SCREENING ATTENDANCE-TWIN APPROACH 30
FIGURE 29 - CLASS DIAGRAM FOR SCREENING SCENARIO 32
FIGURE 30 - DESIGN OF MESSAGE (SCREENING) WITH OBJECTS AND RELATIONSHIPS 33
FIGURE 31 - SAMPLE SCREENING REPORT SENT TO GP IN HARD-COPY FORMAT 34
FIGURE 32 - INTERACTION DIAGRAM FOR THE TWIN MESSAGES 35
FIGURE 33 - BSAMP PROTOCOL ARCHITECTURE 40
FIGURE 34 - SCREEN SHOT OF PREDICTION AND NON-ATTENDANCE DISPLAYED IN FRAMES WITH DIFFERENT SET OF FONT COLOURS 41
FIGURE 35 - SCREEN SHOTS OF PREDICTED NON-ATTENDANCE DISPLAYED AS ‘AMBER’ AND NON-ATTENDANCE AS A ‘RED’ POP-UP PROMPT ON THE WOMEN’S CLINICAL RECORD 42
FIGURE 36 - DATA MINING STREAM FOR AI-ATT ALGORITHM (ADAPTED FROM AROCHENA’S PROJECT REPORT (AROCHENA 2003) 45
FIGURE 37 - AI-ATT ALGORITHM WORKING PROCESS (AROCHENA 2003) 46
FIGURE 38 - PRE-PROCESSING MODULE FOR PROPOSED PROTOTYPE 48
FIGURE 39 - HIGH LEVEL USE CASE DIAGRAM OF THE PROTOTYPE 50
FIGURE 40 - DATA PRE-PROCESSOR MODULE CLASS DIAGRAM 51
FIGURE 41 - AI MODULE (JAABS ALGORITHM IMPLEMENTATION) 54
FIGURE 42 - JAABS ALGORITHM (CLASS DIAGRAM) 55
FIGURE 43 - JAABS ALGORITHM 56
FIGURE 44 - THE ROC CHARACTERISTICS FOR THE JAABS ALGORITHM 57
FIGURE 45 - ROC VALUES ON THE PREDICTED WOMEN (IN NUMBERS) FOR BPNN AND RBFN NETWORK 58
FIGURE 46 - TEST RUN FOR CR TEMPLATE VERIFICATION 61
FIGURE 47 - SELECTION CRITERIA USED FOR THE ENTIRE DATASET 61
FIGURE 48 - FINAL VERSION OF CR TEMPLATE 62
FIGURE 49 - PIE-CHART SHOWING RECORDS CLASSIFICATION AFTER PRE-PROCESSING 63
FIGURE 50 - TRI-FURCATION OF DATASET INTO TRAINING, VALIDATION AND TESTING SET FOR SEVEN EPISODES 64
FIGURE 51 - RBFN STREAM PREDICTION DETAILS FOR ALL EPISODES 70
FIGURE 52 - PREDICTION CHARACTERISTICS FOR ANNP 72
FIGURE 53 - JAABS VS AI-ATT AVERAGE ROC VALUES FOR FIRST FOUR EPISODES 75
FIGURE 54 - CONSULTATION RATE PER PERSON YEAR TREND (1995-2006) FOR AGE GROUP BETWEEN 50 AND 69 83
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE 1</td>
<td>EXTRAPOLATION OF CANCER NOT DETECTED DUE TO NON-ATTENDANCE</td>
<td>6</td>
</tr>
<tr>
<td>TABLE 4</td>
<td>PRE-SCREENING INTERACTION TABLE (PRSC_IN060001)</td>
<td>36</td>
</tr>
<tr>
<td>TABLE 5</td>
<td>POST-SCREENING INTERACTION TABLE (PRSC_In060002)</td>
<td>36</td>
</tr>
<tr>
<td>TABLE 6</td>
<td>CRYSTAL REPORT TEMPLATE FOR DATA EXTRACTION</td>
<td>51</td>
</tr>
<tr>
<td>TABLE 7</td>
<td>JAABS ALGORITHM’S ROC METRICS</td>
<td>57</td>
</tr>
<tr>
<td>TABLE 8</td>
<td>RECORDS BREAKDOWN DETAILS AFTER DATA PRE-PROCESSOR MODULE</td>
<td>62</td>
</tr>
<tr>
<td>TABLE 9</td>
<td>TRAINING, VALIDATION AND TEST SETS CLASSIFIED BY THE PRE-PROCESSOR MODULE</td>
<td>63</td>
</tr>
<tr>
<td>TABLE 10</td>
<td>PREDICTOR ATTRIBUTES FOR AI-ATT ALGORITHM</td>
<td>65</td>
</tr>
<tr>
<td>TABLE 11</td>
<td>VARIABLES ASSOCIATION LEVEL WITH SCREENING ATTENDANCE FOR THE NEW DATASET</td>
<td>66</td>
</tr>
<tr>
<td>TABLE 12</td>
<td>AVERAGE ROC CHARACTERISTICS FOR JAABS WITH AI-ATT VARIABLES</td>
<td>67</td>
</tr>
<tr>
<td>TABLE 13</td>
<td>SCREENING DISTANCE TEST ROC VALUES</td>
<td>67</td>
</tr>
<tr>
<td>TABLE 14</td>
<td>ROC CHARACTERISTICS FOR JAABS ALGORITHM</td>
<td>73</td>
</tr>
<tr>
<td>TABLE 15</td>
<td>ROC CHARACTERISTICS OF AI-ATT ALGORITHM</td>
<td>74</td>
</tr>
<tr>
<td>TABLE 16</td>
<td>ISSUES COVERED BY THE QUESTIONNAIRE (BASKARAN ET AL. 2006)</td>
<td>78</td>
</tr>
<tr>
<td>TABLE 17</td>
<td>DICHOTOMISATION OF THE VARIABLES CATEGORISED BY QUESTIONNAIRE SUBSECTIONS</td>
<td>79</td>
</tr>
<tr>
<td>TABLE 18</td>
<td>CROSS TAB ANALYSES ON SCREENING, PREDICTION, EDI AND INTERVENTION AGAINST DEMOGRAPHIC DETAILS</td>
<td>80</td>
</tr>
<tr>
<td>TABLE 19</td>
<td>PROJECT WORK’S ORIGINALITY SUMMARY</td>
<td>92</td>
</tr>
<tr>
<td>TABLE 20</td>
<td>PROBABLE NUMBER OF LIFE SAVED BY JAABS ALGORITHM</td>
<td>93</td>
</tr>
</tbody>
</table>
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)
Introduction

We provide an introduction to the project work and its focus, objectives and deliverables. This project proposes a novel approach for leveraging knowledge through prediction in a multi-tier architecture. The knowledge leveraging addresses the breast screening non-attendance and its related issues. This project work is aimed at providing the first step towards a multi-pronged strategy envisaged to deliver better primary and secondary care from breast screening perspective. This strategy is also the first of its kind in the UK that addresses breast screening non-attendance from a Knowledge Management perspective.

Setting the scene

Breast cancer is the most common cancer in women with over forty thousand women being diagnosed with the disease each year in the UK (Cancer Research UK 2005). Any information related to the breast can largely affect a women’s consciousness and a threat of breast cancer will have varying impacts on women’s psychology (Cassandra 2006). The incidence of breast cancer is greater within higher social classes resulting in increased awareness of breast cancer and its prevention (Chris et al. 2003). Typically breast cancerous cells originate in the mammary glands (lobules) or in the ducts connected to these glands (Figure 1) or in other tissues around these glands (American Cancer Society 2005). When in close proximity to the lymphatic system, these cells can result in being carried to other organs of the body. This subsequently results in cancerous growth in that organ and is described as metastatic breast cancer (American Cancer Society 2005). Although many causes had been identified of breast cancer, the knowledge of finding a cure is still not within the reach of modern medicine.
Breast cancer should ideally be diagnosed at the earlier stages of its development. Possible treatments include removing or destroying the cancer cells to avoid the spread of the affected cells. Breast self examination (BSE) is an effective and non-intrusive type of self diagnosis exercise for checking any abnormalities/lumps in the breast tissue. Unfortunately this greatly depends on the size of the lump, technique and experience in carrying out a self examination by the woman (Oikonomou et al. 2003). An ultrasound test using sound waves can be used to detect lumps but this is usually suited for women aged below thirty-five owing to the higher density of breast tissue (American Cancer Society 2005) .. Having a tissue biopsy via a fine needle aspiration or an excision is often used to test the cells for cancer. These tests are mostly employed in treatments or post-treatment examination and as second rung diagnostic confirmation methods (Marcela 2004). Performing a Computed Tomography (CT) or a Magnetic Resonance Imaging (MRI) scan would result in a thorough examination of the breast tissue but this technique is not favoured due to reasons including that it may not be economical, needs preparation, is noisy, time consuming and the images may not be clear (Marcela 2004).

Mammography is a technique for detecting breast tissue lumps using a low dosage of X-ray. This technique can even detect a three millimetre sized lump. The X-ray image of the breast tissue is captured and the image is thoroughly read by experienced radiologists and specialist mammogram readers (Marcela 2004). Preliminary research suggests that women aged fifty five and above are more susceptible to breast cancer; mammography is more suited to the women aged fifty five and above (due to the lower density of breast tissue) (Blanks et al. 2000) .. Even though mammography has its critics - mainly due to its high rate of false positives and false negatives (Epstein 1979, Burton 1997) - it has still become the standard procedure for screening women by the National Health Service (NHS) National Breast Screening Programme UK (Forrest 1986). Mammography is the best and most viable tool for mass screening to detect cancer in the breast at an early stage (Medicine net 2002) however, the effectiveness of diagnosis through screening is directly dependent on the percentage of women attending the screening programme (Pirjo et al. 2001).

The NHS Breast Screening Programme, catering to the entire eligible female population is funded by the Department of Health in the UK and is the first of its kind in the world. It covers nearly four million women and detected more than thirteen thousand cancers in the screened population for the year 2005 (NHS review 2006). Currently the screening programme routinely screens women between the ages of fifty and seventy, and employs two views (Figure 2) of the breast, medio-lateral and cranio-caudal (Department of radiology, Brigham and women’s hospital 2007).
Breast Screening Programme in the United Kingdom

High mortality and morbidity rates due to breast cancer have sustained the need for a continuous impetus on research and strategies focussed on preventive medicine (Turner et al. 1994, Majeed et al. 1997, Sin, and Leger 1999). Breast cancer is still one of the major cancer killers in women. The lack of breakthroughs in finding a definitive cure means that preventive medicine is the only viable alternative in reducing deaths due to breast cancer (Fox et al. 1991). Various randomised controlled trials (RCTs) in Europe and around the world have concluded that screening can be an effective tool in fighting breast cancer (Bjurstam et al. 1997, Majeed et al. 1997, Duffy et al. 2003). Technology breakthroughs in imaging and digitization have increased the efficacy of mammograms and countries such as the UK, Sweden, Australia, Canada and the USA all have well-established breast screening programmes (Turner et al. 1994, Majeed et al. 1997, Sin and Leger 1999, Taylor et al. 2003).

The UK National Health Service (NHS) National Breast Screening Programme (NBSP) is unique as it provides free breast screening for the female population aged between 50 and 70 at a national level (Cancer Research UK 2004, Forrest 1986, Tabar et al. 1985).

The recent increase of the upper age limit from 63 to 70 for screening and making a two-view mammogram mandatory has greatly increased the efficiency of benign or malignant tumour detection (Patnick 2006). The NBSP currently runs a massive screening programme catering to almost two million eligible women across the UK (Cancer Research UK 2004). This programme runs on a call/recall cycle which screens all eligible women in a three year interval. The information published by the UK Government Statistical Service (NHS Health and Social Care Information Centre) in its Community Health Statistics report for the year 2006 agrees that, for the past ten years since 1995, the uptake has remained constant at around 75% (Protti 2005) (Figure 3).
From 2002, women aged 65 to 70 had an uptake of 70%. In addition to this, the report also indicated that the number of cancers detected steadily increased from over 5000 in 1995 to nearly 12000 in 2005 – refer Figure 4 (NHSBSP 1999-2005). Deaths due to cancer have gradually decreased from over 11,000 in 1995 to approximately 10,000 (Figure 5). This not only suggests that the number of clinically detected breast cancers has increased, but also confirms that deaths due to breast cancer have steadily declined over the years (using the data until 2004, NHSBSP 1999-2005).
All these inferences indicate that mammogram-based screening initiatives are valuable and do contribute to reduce breast cancer related deaths. The question is whether the screening programme is working efficiently. The efficiency can be mapped to the screening attendance. Figure 3, Figure 4 and Figure 5 clearly reveals that the most important inference that can be deduced from these figures is that the number of non-attendees has been significantly increasing and has reached half a million.

Simple projection of this data, through translation (Table 1), submits that nearly 4,000 cancer incidences would not have been diagnosed. Even if a small percentage of these non-attendees could be made to attend, it would result in the saving of significant lives. Indirectly it can also be inferred that, despite focussed efforts on these non-attendees for the past ten years, there was no real effect on their attendance. Moreover, early stage cancer detection would have a huge impact in reducing cancer related deaths (Baskaran et al. 2006a, 2006b). From these facts and data, it can be confirmed that the primary concern is to reduce non-attendance (Bankhead et al. 2001).
Challenges being addressed

Detecting cancer at its early stages is the only viable strategy to decrease cancer related mortality. This decisive aspect has driven the urgent need to focus more on screening attendance. Maximising screening attendance can only be achieved through a well coordinated delivery of better healthcare (healthcare) to the population. Based on reports by two Swedish county trials (Bjurstam et al. 1997, Duffy et al. 2003) and subsequent recommendations by Forest in 1986, a 70% attendance was proposed as a benchmark for future coverage statistics. In spite of various efforts focussed on this issue, the attendance rate has not improved since this benchmark metric. Even though the 70% screening attendance initially seemed to be a challenging statistic to be achieved, it was never improved upon over the years. The current UK national attendance target for breast screening is 80%. The increased screening population can be attributed to a variety of reasons including the post second world war population explosion, “baby boomers” and the general longevity of human life in this century that has forced health strategists to rethink existing health policies. This has created the opportunity for continuous research, aimed at different facets of care delivery. One such area was the age limit prescribed for the breast screening programme.

Such continuous research had indicated that it is a viable strategy to increase the upper age limit from 64 to 70, thus providing extended mandatory care delivery periods (DOH 2006). According to the Warwickshire, Solihull and Coventry Breast Screening (WSCBS) service manager; the National Screening Service is seriously considering to increase the age band at both ends by an additional three years (breast screening process would start when women are aged 47 years and continue until they are 73 years). This would effectively result in the increase of the eligible women population who would be participating in the future NBSP.
As indicated earlier, the necessity for increasing the quality of care services has prompted the NHS Cancer Screening Programmes to offer a “two view” mammogram (in 2003) as a way to improve the efficacy of detecting screening abnormalities (Patnick 2004, Osborna et al. 2006). Coupled to this, the initiative to gradually digitise mammograms had called for increased investments and has motivated screening units across the country to take advantage of the new technologies available. The digitization of mammograms has not only improved the quality of images but has also facilitated the access, storage (archiving) and sharing of images across geographical boundaries enabling better utilisation of image readers’ (Pisano et al. 2002, Skaane et al. 2003, Skaane and Skjennald 2004). Such mentioned care enhancements are putting undue pressure on the NHS Cancer Screening Programme which is already overstretched in terms of resources. The problems became more complex when such resources are underutilised due to static, below-par attendance rate. Proactive and acute healthcare are vital for cancer control and cure; this has reinforced the necessity to improve screening attendance.

Solution proposed

The above-mentioned challenges can only be addressed by a resource-saving strategy which has better healthcare at its core. As part of this strategy Knowledge Management (KM) was identified as the holistic tool. Techniques associated with knowledge creation and sharing were viewed as approaches for better resource utilisation. Based on the above specifics, an automated software solution - coupled with a humanistic strategy - was designed. The first component in the proposed strategy is related to knowledge creation via Artificial Intelligence (AI), employing Neural Network (NN) algorithms. This strategy also includes a Service Oriented Architecture (SOA) to deliver the envisaged knowledge sharing as the second component.

This research proposes to unify the existing National Breast Screening Computer System (NBSS) software onto a single platform and create prototype software components based on Open Source technologies. The proposed prototype software would be automated to produce the pre-processed data and eventually normalise the data for AI (neural network) assimilation. These activities would be performed sequentially without human involvement for repeatability, reliability and accuracy. The AI model for attendance prediction itself would be simulated on the Open Source technology platform. This model incorporates all additional transformations occurring within the screening process (including the change in the screening upper age limit). The prototype framework proposed will incorporate the AI model for knowledge creation (i.e. list of predicted non-attending women). The prototype combines the demographic data pertaining to the non-attending women and information related to her General Physician (GP) as a messaging package. This package triggers the generation of an electronic message based on the Health Level 7 (HL7) version 3 standards and utilises Service Oriented Architecture (SOA) as the message delivering technology (refer to Figure 6).
Research Question

Earlier research in identifying the factors contributing to non-attendance prediction was tested in the WSCBS unit (Arochena 2003). This research was the first step in identifying attributes that can predict women’s attendance a priori. This prediction algorithm was developed by an expert statistician through visual modelling software in an academic setting and there is a lack of associated research as to how this prediction can be put to effective use. Over the last decade, international research pursued by Melville et al. 1993, Bryant 1996, Richards et al. 2001 and Zorbas 2003 had indicated without doubt that primary care can play a vital role in alleviating the problem of non-attendance through interventions (Bankhead et al. 2001, Brunton and Thomas 2002).

My primary research question is:

*Can the breast screening attendance be increased through appropriate Knowledge Management tools and techniques?*

The hypotheses this project sets out to validate are:

- that it is possible to design and implement an Open Source-based prediction algorithm prototype for breast screening attendance
that it is possible to design a messaging platform for knowledge creation (through prediction) and sharing model (knowledge/information exchange architecture) with available technology and infrastructure

that it is feasible for GPs to make interventions to increase breast screening attendance through electronic prompts

Research Objectives

This project proposes to integrate the existing call / recall process of the Breast Screening Programme (BSP) with the results obtained by a newly developed predictive algorithm in order to increase uptake in the screening programme. This will be achieved through the use of Knowledge management (KM)-based tools and techniques. The developed protocol will be the first step in providing the ability to target non-attendance at an early stage and initiate counter measures automatically.

The objectives included:

• To investigate and evaluate current practices in the BSP call / recall process

• To investigate and evaluate the suitability of the existing Electronic Data Interchange For Administration, Commerce and Transport (EDIFACT) and Health Level 7 (HL7) standards

• To design and analyse a questionnaire to evaluate the GPs' role in reducing BSP non-attendance

• To identify a messaging platform and a viable architecture for integrating the Java based Attendance prediction by Artificial Intelligence for Breast Screening (JAABS) algorithm

• To develop a Knowledge Management protocol based on the above, followed by the implementation of a prototype (JAABS algorithm) encompassing the required standard, ready for knowledge sharing with healthcare stakeholders via the NHS' National Care Records Services (NCRS) infrastructure.

The primary deliverables of this work are a set of breast screening messaging protocols which meet the standards set by Health Level 7 (HL7) UK and implemented via tools recommended by HL7 and the National Health Service (NHS) Connecting for Health (CfH). The work detailed in this project builds on an earlier research conducted within the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group at Coventry University. The information and knowledge sharing framework that has been proposed for deployment on a Service Oriented Architecture (SOA) operates on web services-based technology and draws parallel to the UK government’s e-gif policy (Ward 2003). The prototype proposed would be the first software iteration in building a real-time functioning KM-based system which would enable a seamless transfer of knowledge for better healthcare delivery involving screening unit and primary care (GP).
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org
The project deliverables are:

- an evaluation report on the suitability of the HL7 Version 3.0-based XML standards would be conducted for its appropriateness to the proposed BSP Messaging (BSPM) protocol

- an extension of the evaluation report to compare the proposed HL7 Version 3.0 with EDIFACT based protocol for highlighting salient features

- a framework for implementing the JAABS algorithm

- a prototype implementation to predict non-attendance by the JAABS algorithm

- tests of the execution, evaluation of the prototype with real-time dataset

- a design for the proposed BSPM messaging protocol will be designed

- an architecture for the proposed prototype

- a questionnaire-based evaluation of the GPs’ role in reducing the BSP non-attendance and justification of the proposed prototype’s KM based strategy

- technical reports

- dissemination of findings would be via academic presentations, paper publications and conferences
Overview

The current research can be visualised as a complete set of proposed protocols within the healthcare domain. The core component, namely Knowledge Management (KM), Artificial Intelligence (AI), and Health Informatics (HI), brings the necessary synergy into the research and transforms it into truly inter-disciplinary work (Figure 7). This invigorates the breast screening domain and provides the impetus required for new approaches.

Figure 7 – Chapters’ relationship with the research core areas

The specific components of the current research and their relationship to the research chapters are shown in Figure 7. The diagram appropriately positions the chapter with reference to their content and its association to the core areas of the research. Some chapters are repeated in different parts of the diagram; that signifies a multidisciplinary coverage of the specific chapter.

We discuss the introduction, challenges being addressed and the research objectives. Chapter 2 (which has been removed for the purposes of this document) details the literature review and a summary of the research areas related to this study namely Knowledge Management, Health Informatics and Artificial Intelligence. As part of the messaging standard evaluation, a comparison between EDIFACT and HL7 is included at the end of this chapter. We provide the details of knowledge, its creation and sharing. This is followed by a framework which highlights the human-centric nature of knowledge and a methodology based on KM (specifically for healthcare domain) is proposed. This chapter concludes with a detailed description of the proposed Breast Screening Attendance Messaging Protocol (BSAMP) and substantiates the importance of HI in healthcare knowledge and information dissemination.
We go on to deal with the AI component of this project which incorporates data processing and the proposed JAABS algorithm and its evaluation with the earlier algorithm Artificial Intelligence based ATTendance predicting algorithm (AI-ATT). This chapter also enumerates the design and implementation logic employed by the JAABS algorithm. An evaluation and testing report of the JAABS algorithm concludes this chapter. Chapter 5 identifies the GPs’ role for the knowledge assimilation and the types of intervention that can be initiated to increase the screening attendance. A comprehensive analysis of the questionnaire on GP intervention and results are also discussed in this chapter.

We conclude this project and highlights the changes necessary for adopting the proposed BSAMP protocol. This project also lists the conceptual developments accomplished in a separate subsection and highlights how the objectives were not only achieved but, in many cases, were exceeded. This chapter shows the full richness of this project and reflects the new concepts added within the fields of healthcare informatics and Knowledge Management and how these new concepts will further the knowledge boundaries. The last subsection details the challenge faced and lays out future work for continued research in this domain.
Proposed Knowledge Exchange

This research illustrates the importance of Knowledge Management (KM) and its effective usage in healthcare domain. Newcomers to KM, especially healthcare specialists, are often overwhelmed when trying to understand the intricacies of KM in healthcare. These challenges provided the necessary impetus to the author for investigating the various KM concepts and elucidate its principle components with clarity for easier understanding.

The core for KM is knowledge and its management of its various avatars. It is fundamentally important to understand and define knowledge from the KM perspective. The following sections deals with the various views and suggestions from the author on KM, starting with a definition of knowledge from KM context.
Information and knowledge

A misconception exists in differentiating knowledge from information that has attracted a continued debate on the legitimacy of the KM paradigm itself (Wilson 2002). The schematic depicted in Figure 8 aims at differentiating information and knowledge within a healthcare perspective. The figure's analogy of data can be interpreted as a set of raw, unprocessed numbers (75, 77, 57, 70, 72); contextual details when added to this set of data - such as (1) this dataset represents heart beats per minute and (2) pertaining to male aged 62 years - can add lucidity to the dataset provided, becoming information (Miller 2002). When a heart specialist assimilates this information, a unique process (which is an inherent characteristic of the human brain) of identifying connections, similarities and patterns in the information provided using his/her earlier insight on the subject initiate new knowledge about the patient and may result in the conclusion of symptoms of Arrhythmia (Koskinen 2003). The said steps constitute the initial phase of the knowledge cycle.

Such knowledge cycles will contribute further knowledge creation and result in adding clarity and assist in achieving wisdom. As a routine process of sharing this acquired knowledge, an individual has to resolve this issue through traditional knowledge sharing techniques such as writing notes, manuals, reports and books coupled with lectures, meetings and the like. The knowledge residing in the human brain is termed Tacit (or implicit) knowledge and the other externalised forms of it are termed Explicit (or expressed) knowledge. There is a grey area in differentiating the explicit knowledge from information. Under some circumstances, these can become interchangeable but the human context alone differentiates the information from explicit knowledge (Augier et al. 2001). In expressing his/her tacit knowledge, the heart specialist resorts to the creation of notes based on the patient's information. This form of knowledge, when viewed by any individual from the heart specialist's perspective, is termed explicit knowledge.

When the same notes are referred to, by a neurosurgeon for cerebro-vascular accident diagnosis, the explicit knowledge created through these same notes becomes information (since the current interpretation of the notes has lost its creation context). Therefore the expressed knowledge viewed in the knowledge creator's context is explicit knowledge, but not otherwise (Rodrigo 2001). This completes the knowledge cycle which can be dichotomised into distinct sets of activities, one related to knowledge creation and the other to knowledge sharing. Such knowledge cycles overlap and feed on each other in propagating knowledge through the time continuum.

Knowledge creation

The dichotomisation of activities assisted in addressing the issues related to Knowledge Creation (KC) and Knowledge Sharing (KS). Tacit knowledge creation within a human brain can be viewed as a combination of input stimuli, even though there have been different foresights in explaining how knowledge is created in cognitive environment from a biological and psychological sciences point of view (Binnie and Williams 2002). This discussion would limit itself to the cognitive side of knowledge creation as a black box and rather focus on the inputs and outputs of human cognition (Schwenk 1988, Thomas et al. 2001). In general the inputs to knowledge creation can be viewed as a set of stimuli which
can be classified into four main groups: explicit knowledge, interaction, information, and insight – which coalesce in a continuous spiral process (Figure 9) to create new knowledge (Baskaran et al. 2004). The external stimuli received by the human brain through its various sensory organs can be in the form of explicit knowledge where the knowledge consumer views the knowledge from the creator’s context (Figure 8). This could range from physician’s notes to a financial analysis report. The interaction stimuli are related to learning through physical experience.

![Figure 9 - Knowledge creation spiral](image)

The physical experience can be an articulated exchange of knowledge with an expert or could be gained through physical experience such as learning to drive a car through heuristics. Information stimuli are related to knowledge creation through the 1st G or 2nd G information (Figure 8). This stimulus depends on the knowledge receiver’s ability to comprehend the information using (self) past experience. Insight originates within; hence it could be viewed as an internal stimulus. This cognition-based stimulus can be attributed to all previous information, knowledge and wisdom possessed by the individual. It also has a genetic aspect of knowledge retention. The above said stimuli in combination or alone create new knowledge (Baskaran et al. 2004).

**Knowledge Management (KM)**

The earlier sections described in detail the different flavours of knowledge and stimuli for its creation in a cognitive environment. This section briefly describes the concept of KM. In contemporary terms, the view of KM can be summed up as “The process of lowering the transaction costs associated with creating, sharing, and applying knowledge, and developing improved strategies to support these activities” (Prusak and Matson 2006). After careful evaluation it can be concluded that humanistic factors can be attributed to the intricacies experienced in present day KM (Prusak 2001, Prusak and Matson 2006, Ichijo and Nonaka 2007). The process of knowledge creation and sharing is inseparable from its human

KM has been construed and viewed in different perspectives, some view as synergy, some as Intellectual Capital (IC), competitiveness, a decision making tool, customer focus strategy, employee relation and development, lowering costs, customer value addition, etc (Chourides et al. 2003, Sveiby 1998). Healthcare organisations have started to believe that long-term profitability and business sustainability are primary concerns and are to be addressed outside the individual healthcare domains. So to achieve this, an organisation-wide KM strategy has to be adopted (Ichijo and Nonaka 2007, Kasvi et al. 2003). This suggests that KM has to be viewed more in terms of leveraging than managing. Limited, objective-based exposure has removed knowledge from a holistic concept and forced business managers to view it as a resource which can be managed (Wernerfelt 1984, Swan et al. 1999). A more proactive approach would refer to the unbounded subjective nature of knowledge and rather conclude that knowledge should be leveraged (Prusak, 2001). Since KM has been in use for over a decade and has been well established in both academic and organizational domains, it will be appropriate to term all knowledge activities under KM, in spite of the fact that KM should focus on leveraging knowledge (Walsham 2001, McDermott 1999).

Knowledge creation is still considered entirely as a cognitive process (Thomas et al. 2001, Prusak and Matson 2006, Bell 1999, Ichijo and Nonaka, 2007, Pemberton et al. 2001, Baumard 1999). This process has not been completely understood by modern day biological science and hence forces knowledge workers to look at it as a black box. From an organisation’s perspective, the focus should be on the conducive environment for knowledge sharing to create new knowledge (a knowledge cycle) among individuals within the organisation (Thomas et al. 2001, Baumard 1999). Hence importance has to be given to knowledge cycles, thereby assisting in the transcending process of creating a knowledgeable organisation. KM oriented activities, strategies and processes aim at achieving a favourable environment that implicitly propagates productive knowledge cycles.

The objective nature of knowledge was seen as an opportunity by many KM solution providers. In the 1990’s such KM solution providers exploited organisations which showed interest to invest in their readymade technological based KM solutions (Prusak and Matson 2006, Walsham 2001, McDermott 1999). Irrespective of the huge investments incurred in implementing such technology-based KM solutions they utterly failed to deliver the much promised returns on such huge investments. This resulted in a negative view on KM; progressively organisations were wary about committing to future KM initiatives (Walsham 2001, McDermott 1999). This caused overexposure on non-core KM issues, which deprived KM of the much needed maturing time. This misunderstanding can only be corrected by refocusing KM on its precise core issues i.e. humanistic aspect. Liebowitz et al. (2005) acknowledges the importance of people and culture as 80% in successful KM and attributes the remaining to technology. In spite of the problems there is a growing interest in KM, particularly in healthcare organisations. They have started to realise the potential of KM by concentrating on the humanistic attributes of knowledge creation and sharing (Davenport 1995).
KM has been accepted as a strategic management science (Drucker 1993, Argyris 1992, Senge 1990). Especially in organisations where exigencies as mentioned in the previous section are profuse, KM is looked upon as the antiphon (Disterer 2002). Organisations have recognised both sticky and leaky features exhibited by knowledge (Brown and Duguid, 2001). Earlier sections enumerated the nature of knowledge and the incidence of it can only be manifested through humanistic traits (Ichijo and Nonaka 2007, Walsham 2001). In addition to this, any organisation, especially healthcare is identified as people centric. This underlines the importance of human nature within knowledge in organisations (Walsham 2001, Chourides et al. 2003). Every individual’s physiognomy has its own complexion- rules, morality and economic associations, and this renders a unique phenomenon attributed to that particular individual. Any organisation can be viewed as a layered model from the knowledge perspective (Figure 10) (Ichijo and Nonaka 2007).  

![Figure 10 - Multiple facets of Knowledge model in Project-based organisation](image)

The core of the model is human psychology, as mentioned earlier all knowledge originates at this level. Knowledge creation also takes place at this level. Human psychology either acts as a knowledge enabler or a deterrent. Human values and beliefs are important factors for urging the individual to share knowledge withheld in the individual’s cognition. The stickiness manifests in this layer, this can be attributed to the cognitive facet of knowledge creation. The knowledge leakiness can also be associated to this layer as it is related to individuals leaving the organisation. At this level, Maslow’s theory of human needs can be an ideal mechanism to explain and draw parallels regarding the motivating factors for emphasizing knowledge creation and sharing (Maslow 1943). There are different theories that explain factors which have a strong relationship in explaining an individual’s view on knowledge sharing with other individuals from a personal perspective (human psychology layer- refer to Figure 10). Personal level ego based human attributes influence the
individual’s intent to share knowledge (Durkheim 1958). These factors are not permanent and may fluctuate based on the environment to which the individual is exposed and its combinatorial effect on the individual’s cognitive framework.

The second layer in the model is social psychology (Figure 10); this plays a major role in sharing knowledge and in creating knowledge at group level. Some knowledge theorists concur that “the sum is greater than the individuals” (Bresnen et al. 2003). Any organisational team is a permanent/semi-permanent or temporary agglomeration of individuals with a common purpose of achieving organisational objectives in time and to deliver the set goals within stipulated resources; such organisational structure creates a social bond among the team members. Even though sociological aspects can only be deduced from a psychological perspective, a collective sense can stem from human’s social nature.

Epistemic barriers can arise due to the complexities existing within a formal organisation. Polanyi argues that social component is mandatory in all forms of knowledge creation and sharing (Polanyi 1966). This concept of belonging and inter-responsibility can have an adverse impact on how a group the team will share the acquired knowledge not only created within the organisation’s temporal continuum but also knowledge already possessed by the individuals. The individual’s cultural, ethnic and education background affects the intensity with which the individual will share knowledge (Fernie et al. 2003, Brown and Duguid 2001). The individual’s inclination and eagerness to share knowledge can be dampened by the current organisational culture. It can aid or deter the individual by creating social apprehensions, peer pressure and aloofness among the group (Fernie et al. 2003, Brown and Duguid 2001, Newell and Huang 2005).

The concept of “Communities of Practice” (COP) is mostly constituted among peers who are equal in stature in the organisation; this does have its inherent advantages such as uninhibited knowledge sharing through cross-fertilization of ideas (Brown and Duguid 2001). Individuals pool their knowledge resources and act as temporary think-tanks within the organisation; this can be mapped to the social psychology layer (Figure 10). This type of COPs can be sustained beyond individual activities for which these COPs were constituted within the organisation; thereby creating a forum which can be looked up for finding solutions which are difficult from an individual’s capacity (Newell and Huang 2005, DeFillippi 2001). These groups would put the organisations betterment over individual’s gain. Experts concur that lack of proper recognition and appreciation will abrogate the “stickiness” of knowledge, but on the contrary, this will facilitate the generation of knowledge-islands in organisational environs. Incessant support to such knowledge-islands creation would ultimately result in ineffectiveness of the organisation’s functional group due to a decline in innovation and creativeness. The stickiness can only be abated by supportive strategies addressing the social and communal apprehensions rather than the individual’s beliefs (Hall and Sapsed 2005). This discussion would take the concept of social groups such as COPs to the next level by expanding them to play a bigger role in sharing knowledge irrespective of the individual’s position in the organisation’s domain (Hall and Sapsed 2005). All the organisational staff should be forthcoming in sharing their knowledge at the right time which can make a huge difference in the organisation’s success or failure. The knowledge sharing culture should be viewed as a holistic characteristic of the team and
the organisation has to cultivate such an environment which can thrive and sustain a cordial atmosphere for expansive knowledge sharing.

The next layer in the model (Figure 10) refers to the organisation’s cultural environment which is vital for creating an enabling atmosphere for knowledge creation and sharing (Lytras and Pouloudi 2003, Stonehouse and Pemberton 1999). This atmosphere should cater to the psychological stratum at both the personal and the social level (Fernie et al. 2003). A strong and dedicated leadership and top management which has self-belief in knowledge creation and sharing as an intrinsic organisational activity are mandatory (Marr 2004). Only dedicated management which sets itself as role model can propagate a knowledge enabling culture. Moral support and encouragement originating from the top management alone can sustain a knowledge environment throughout the organisation (Marr 2004). Durkheim in his book on social theories postulates that when human subjects are intentionally or unintentionally made to follow a routine and repetitive enforcement of constraints this would gradually give rise to habits and internal tendencies which render the constraints unnecessary in future (Durkheim 1958). Similarly, routine knowledge creating and sharing attitudes can be ingrained within the organisation’s work ethics. This would not only ensure a holistic application of knowledge but also enable new individuals joining the organisation to be overwhelmed by the positive culture towards knowledge and embrace it wholeheartedly, which results in the inhibition of psychological aspects in the individual and a social approach alone will be gleaned out (Durkheim 1958).

Through these features, the organisation can realise its potential related to the beliefs, tendencies and practices of its members. This eventually culminates into a truly knowledge-based social phenomenon. The organisation’s infrastructure has to be viewed as an enabler for the knowledge environment (Stonehouse and Pemberton 1999). Many simple tools such as an e-notice board, e-fora, blogs, discussion boards, mailing system, and a database for useful information are vital (Wernerfelt 1984, Swan et al. 1999, McDermott 1999, Liebowitz et al. 2005). Providing access to facilities for knowledge retrieval and sharing are fundamental at this level, but every organisation has to realise that technology can play only a supporting role and cannot replace the humanistic core of KM (Wernerfelt 1984, Tissen et al. 2000, Liebowitz et al. 2005). If a healthcare organisation can focus on the aforementioned aspects of a knowledge-based organisation model, it can harvest many positive outcomes. The outcomes are not only limited to a better execution of activities within the organisation, but also increase the values of the organisation’s different facets such as Intellectual Capital (IC), Social Capital (SC) and Knowledge Capital (Tissen et al. 2000, Prusak 2001, Marr 2004). Knowledge Capital is related to both forms of knowledge i.e. tacit and explicit, including the 1st and 2nd G information. These capitals are fundamental for a successful, modern organisation (Edvinsson 1997, Fong et al. 2005). All the above mentioned strategies can work in unison to find solutions for the exigencies mentioned earlier.

**Beyond KM**

Two conflicting perspectives were discussed by experts, Wernerfelt (1984) views knowledge as a resource that could be managed as any other in an organization. Wernerfelt (1984) further details that knowledge exists as an independent entity from its creator and creation context. Where as, Berger and Luckmann (1967) view knowledge as a culmination of shared
belief, prospering in social domain and extensive social interaction enhances knowledge creation (Cicmil et al. 2005). This discussion draws its strength from Berger and Luckmann’s views and expands and details the importance of humanistic core of knowledge. SECI framework (Socialization, Externalization, Combination and Internalization) of Nonaka and Takeuchi (1995) is acclaimed for its focus on socialization as a tool for creating and sharing knowledge (Ichijo and Nonaka 2007).

Some of the challenges which are to be addressed in organisational environs are as follows (Ichijo and Nonaka, 2007):

- Knowledge crossing boundaries between team members of different disciplines
- Invisible boundaries between hierarchical management layers
- Inter-disciplinary knowledge sharing is stronger than intra-disciplinary
- Group discussion can create new knowledge generation
- Integration of knowledge from different stakeholders
- Encouragement to collective learning and inter-organisational learning

When the various humanistic knowledge aspects mentioned earlier are addressed appropriately, any organisation can evolve into a knowledge-based organisation. As solutions, many initiatives such as rewards, recognition, employing knowledge gate keepers/enablers have been suggested (Fong et al. 2005, Cicmil et al. 2005). Organisations and academia had accepted that rewards and recognition would encourage knowledge sharing, but this works only on a short-term basis (Fong et al., 2005). These schemes of appreciation are difficult to measure and benchmark, since the knowledge being shared / created is beyond quantifiable mechanisms due to its rich tacit knowledge content. This may force the organisation’s staff to take advantage of knowledge sharing as a means for personal benefits, rather than for the betterment of the organisation as a whole. This discussion on factors beyond KM, highlights the fact that all knowledge managed staff should come forward to indulge and propagate knowledge sharing eventually creating new knowledge without expecting specific compensation for such acts. They should accept knowledge sharing as a culture and discharge knowledge creation activities as a part of their regular activities. Hence a holistic concept of knowledge sharing and acceptance are imperative for all future knowledge managed organisations.

Some organisations had employed a knowledge worker as a gate keeper of knowledge and they are expected to not only assist in knowledge creation but also enable the free flow of knowledge across organisational boundaries (Walsham 2001, Bresnen et al. 2003, Marr, 2004). But this approach is only intended as a short term arrangement and ultimately all organisations have to embrace knowledge and do sharing as a routine process. A futuristic knowledge sharing can be envisaged as in Figure 11. After the extended use of knowledge workers in expediting their knowledge-based activities they themselves can become obstacles for the free flow of knowledge.
In addition, they will become liabilities to a knowledge system (Bresnen et al. 2003). Their own personal egos, moods and prejudice can hamper the free flow of knowledge (Lytras and Pouloudi 2003). To overcome the above-mentioned shortcomings, knowledge workers have to be replaced by the entire team (executing a knowledge worker’s function as a group) and thereby the organisation can realise an unrestricted flow of knowledge paths not only at the intra level but also at the inter-organisation level. This can easily be expanded into organisation wide knowledge flow paths and even parent organisation’s conglomerates and partners can involve themselves actively in the organisation’s free knowledge sharing (Cicmil et al. 2005). Healthcare Organisations should not only commence KM based initiatives but should be focussing more on sustaining such initiatives for a longer timeframe. This allows the whole organisation to address new challenges by creating new knowledge and sharing them as part of their routine activities. That is when one can realise a complete knowledge focussed healthcare organisation. Care has to be exercised in addressing all the facets of the knowledge model and focus at all levels of the model’s layer to maximise benefits (Figure 11). Many KM models and frameworks have been proposed by experts specifically for healthcare-based organisations (Whelton et al. 2002, Prencipe et al. 2005). The current discussion highlights that such organisations have to select the best framework for their environment; they may even employ a hybrid of more than one framework for maximum efficiency (Whelton et al. 2002, Prencipe et al. 2005). Irrespective of what model/framework is selected, due focus on humanistic aspects has to be dealt at the core level for achieving the maximum efficacy of KM. The model shown in Figure 11 could be used as a guide to orient such KM models on humanistic aspects.
Total Knowledge Management in healthcare (TKMh)

Current KM strategies in healthcare are myopically focussed on how to convert knowledge from a tacit to an explicit state. The Total Knowledge Management in healthcare (TKMh) model, although abstract in nature, emphasises that KM-based implementations in healthcare should be oriented towards the key issue of sharing knowledge in a tacit-to-tacit mode. TKMh highlights the necessity of a human-centric approach to be adapted while implementing KM strategy in healthcare. The model can be best visualised as a series of healthcare management phases (as depicted in Figure 12). The "Initiate" phase brings about a change in the environment to cultivate healthcare personnel's abilities and create tacit knowledge impulsively as a routine healthcare-related activity. The second stage, "share", sets up the environment so that healthcare specialists have the proactive desire to share knowledge and focus more on tacit-to-tacit knowledge transfers.

![TKMh model diagram]

**Figure 12 - TKMH framework**

The "establish" stage signifies that a culture has to be established to generate a network structure and to sustain an uninhibited sharing of knowledge throughout the organization. This should not be limited to one community, function or problem domain. The "exploit" phase completes the iterating loop and would start when the healthcare organization can garner the returns of the established culture through better realisation of its resources in all
forms and deliver better service to the public. The set of phases generates a continuous feedback loop to its earlier stages and re-strengthens the iterating cycle until each phase has attained its highest efficiency. The author believes that any healthcare organization, while adapting a predefined KM framework, should analyse the model's individual components. This would assist in selecting only the necessary individual components thus aligning the framework to suit that particular healthcare organization's unique way of functioning.

Mapping these components to the TKMh model would provide overall guidance for the KM implementation. This would correctly orient the implementation towards a tacit-to-tacit mode of knowledge sharing. The term TKMh should not be confused with TQM (Total Quality Management) as it is unrelated to this legacy quality concept. TKMh is an abstract model aimed at implementing Knowledge Management in healthcare organizations (Baskaran et al. 2004).

Knowledge creation and application of TKMh model for this project

Earlier sections provided the necessary focus on KC and sharing. This section applies the generic KM concepts to the breast screening domain. In doing so evaluates and validates the proposed KM concepts. As mentioned in section, various stimuli are mandatory in order to instigate KC. As KC is fundamentally a cognitive process manifested within the human mind, a favourable environment is required to create new knowledge (Figure 13). This aspect could be best handled only by an organizational strategy that has faith in efficient KM procedures. The change program currently taking place within the NHS is rapidly moving towards this ideal (Baskaran et al. 2004).

![Figure 13 - Knowledge Creation in predicting screening attendance](image)

The NHS Connecting for Health program places emphasis not only on creating the necessary knowledge, but also provides an infrastructure to effortlessly share knowledge within the
NHS’s various clinical pathways (Baskaran et al. 2005). The current research is aimed at automating the process of prediction (the KC process) and efficiently sharing predicted knowledge with primary care deliverers (GPs). Figure 14 depicts the application of TKMh framework for the current research. In the ‘Initiate’ phase, the tacit knowledge creation is commenced by an expert researcher. Research Methodology highlighted the importance of qualitative research and, in this phase, knowledge from the previous researcher was collated by way of semi-structured interviews and open-ended questions. This knowledge is later captured by the AI algorithm for enabling continuous knowledge creation (prediction).

During this phase of knowledge capture, the efficiency of the algorithm was compared with the previous AI algorithm through triangulation. As part of the quantitative analysis, features related to the receiver operator characteristics were also completed. In the ‘Share’ phase, the created knowledge is shared by utilising the NCRS infrastructure through HL7 messages bundles. The ‘Establish’ phase, recognises, creates, and sustains the humanistic values related to the research domain. This phase focuses the primary care deliverers (GPs) to make appropriate interventions. During this phase, the qualitative research aspects were addressed through piloting within which open-ended questions and semi structured interviews were conducted with randomly selected GPs for establishing the nuances of knowledge sharing. This assisted in fine-tuning the proposed knowledge sharing mechanism and helped to ascertain the challenges that should be addressed before this protocol can be fully utilised.
The data collected during this qualitative research stage warranted the need for additional and comprehensive quantitative research (through a questionnaire-based study) to gather as much information for leveraging the proposed KM-based techniques. This would not only accumulate evidence but would also consolidate the healthcare process-related changes necessary for the proposed protocol. This provided an opportunity to triangulate the data collected during the piloting phase interviews with the questionnaire survey findings. The ‘Exploit’, phase would signify the continued efforts by not only primary care but also through other care deliverers to address screening attendance. In addition, the infrastructure can be easily extended to cover not only breast screening attendance but also other national screening programmes. An iterative feedback loop assisted in reinforcing the project ideologies and fine-tuned the research activities.

Knowledge Creation for this research is illustrated in Figure 13. The KC process is instigated by the various forms of stimulus. These stimuli can be mapped to the earlier concepts namely insight, information, interaction and explicit knowledge. The insight stimulus comes from the initial idea of the researcher (the human component) to use automated prediction (prediction analogy) as an effective tool to counter non-attendance. Explicit knowledge acts as a proactive stimulus through knowledge derived from earlier research by the group (on breast screening and AI) and through literature review of various books, journals and academic materials. Information pertaining to the screening history of the women and demographic details were collated from various sources. This constitutes the information stimulus for this research. Interviews and meetings with the BSU’s staff and previous researcher created the necessary interaction stimulus for KC.

The knowledge created by the various stimuli should not be a one time process (where knowledge created would be static). This project work envisages the use of an AI-based Neural Network (ANN) to capture the KC process. In this way, automation can be implemented within the existing breast screening process; where the knowledge created is in the form of screening attendance prediction. The ANN has been implemented using Open Source technologies (Java environment). The implemented ANN works via a Java-based server program. It uses historical screening data and demographical information by way of Townsend deprivation scores (Arochena 2003) as predicting factors (Adams 2004).

This forms the dataset that is presented to input neurons. The middle layer (also known as the hidden layer) is connected to every other input neuron and to one output neuron. The output neuron remains at “Zero” when a woman is predicted as an attendee and turns to “One” when she is predicted as a non-attendee. Earlier research has confirmed that one hidden layer will suffice to map any multivariate type of input domain to the output domain (Arochena 2003). During the training stage, the error function is fed back through the network from the output neuron. The proposed knowledge capture is implemented using the architecture shown in Figure 15. The input layer is presented with different combinations of attributes based on the specific episode on which the network is being trained. This simple back propagation based ANN is adequate for testing the above knowledge capturing event.
The AI-ATT prediction algorithm by Arochena (2003) channels the input through both the pruning network and the RBFN. The final prediction outcome was based on better prediction confidence between the two networks. The input dataset is passed through the corresponding pre-trained network, based on the screening episode number. Extensive network training and testing indicated that the more prediction parameters present (in the form of historical episode details), the better the prediction results. The current research relies on a similar ANN for prediction, it further recommends two approaches to share the created knowledge with the primary care stakeholders (Figure 16). The proposed JAABS algorithm also employs pruning network and RBFN.
Knowledge sharing for this project

Even though KC through proper personification is essential, it would be of little use to an organization unless it is properly mapped to all probable user-interactions within healthcare services. The sharing process should also encourage the healthcare providers to carry out their clinical and ethical responsibilities. Proper recognition (including monetary incentives as short-term catalysts) could be used to encourage continuous improvement of personifying knowledge at the delivery level. The knowledge gained through prediction can best be utilized at its point of origin (breast screening unit). A priori knowledge of predicted non-attendance could probably be used in breast screening units to address empty slots or to arrange standby screening slots or even to use the available equipment and staff’s time suitably.

This knowledge sharing approach within the screening unit has limited scope for improving its efficiency and on the other hand will have little or no progress on its resources front. Additional resource gains can be made when this knowledge is shared in the most efficient way with primary care providers, i.e. GP surgeries. This would provide a unique opportunity for GPs to intervene and increase the screening unit’s uptake (Bankhead et al. 2001). The knowledge sharing process should also address the packaging and secure transmission of knowledge to the surgeries. The GP surgeries may have a better chance of making contact with women even before the screening date. This could instigate a GP intervention. The intervention can be carried out by various methodologies such as:
1. an arranged meeting with the women in person

2. a telephone-based intervention by the GP

3. using an opportunistic intervention when a woman approaches the surgery for some other health-related service

4. to send a letter signed by the woman’s GP explaining the importance of the screening and emphasising the necessity of attending her allotted screening appointment (this letter might be accompanied by more information about breast screening and leaflets in the woman’s native language)

These interventions are aimed at educating woman on breast screening and would cultivate awareness among probable non-attending women thus enabling them to make informed decision. The quantitative studies accomplished through the questionnaire-based survey shed more light on the preferred type of interventions and the suitable healthcare personnel for initiating interventions. This project, as part of the personification process, addresses the challenges encountered at the surgery and recommends possible strategies to make such interventions an absolute success in increasing screening attendance. Only then can the research can be deemed as a successful KM project. Based on this personification strategy the proposed BSAMP protocol can be viewed in two facets. The first facet is related to KC and the second facet is the knowledge sharing (attendance prediction is at the pre-screening and screening results are at post-screening stage). Details of these facets are deliberated in the following section.

**Proposed twin approach**

Earlier sections enumerated the current activities in a screening programme through a set of ‘use case diagrams’. The changes proposed and their relationships with existing activities are shown in Figure 17. This figure also depicts the proposed sequence of actions involved in the KC and knowledge sharing process within a breast screening centre. A marked difference between these set of activities to the earlier ‘use cases’ described in earlier section is the absence of PNL and its related processes. PNL has been discontinued after careful evaluation and research by the National Screening Programmes (Codd, 2006).
This section explains the BSAMP protocol for the KC (phase 1) and KS (phase 2) (Figure 17). Data pre-processing refines the input dataset for ANN utilisation. When the prediction algorithm is presented with the refined dataset, it predicts the non-attendees list. These activities are identified in Figure 17 as phase 1.

**Prediction and post-screening results**

The first approach (Figure 17) refers to the following main actions: the Breast Screening Unit (BSU) manager generates the incoming screening batch. This batching activity is executed on
a three year recursive cycle in such a way that it includes the whole eligible female population covered by the screening unit.

When the batch list is generated by the NBSS software, the manager utilises a report template to export the batch list as a flat file and saves it in a suitable folder located in the BSU server. The next action is to generate the non-attendance list through the automated software. The software component is resident in the BSU server and is accessed by the client program via a Graphical User Interface (GUI). The software prompts the user to point to the location of the flat file (already created by the template and stored in the server hard drive). Once the file is located, the pre-processing software will automatically generate the required predicting factors and normalise the data. The User (via the GUI interface) points the ANN to the location of the historical data (in the flat file) to train the network. Once the training is completed, the net is pointed to the test data so that prediction can be initiated.

Any errors during the pre-processing, training and the actual prediction activities are stored in individual log files which can be viewed at a later point in time for feedback. The GUI gives the option to the user to initiate the Simple Object Access Protocol (SOAP) message. The message body is instantiated with reference to an eXtensible Markup Language (XML) schema definition designed on the Health Level 7 (HL7) version 3.0 standards. The XML message is called upon by the software to generate the SOAP envelope and attaches the XML message to the SOAP body with a digital signature (for security). The Java-based web services technology provides encryption to make the message completely secure. The message is transmitted via web services to the GPs’ mailbox server (Figure 16). Once the GP server connects to the mailbox it downloads the messages and the GP’s software automatically updates the XML content to the women’s records after proper decryption and verification of the digital signature contained in the SOAP message. Meanwhile, the BSU executes its routine process of inviting the women by dispatching an appointment letter (with details of the screening date and time).

For predicted non-attendees, the knowledge of non-attendance generated by the ANN is now available to the GP. When the woman meets the GP for other healthcare related services, GPs can proactively initiate an opportunistic type of intervention. Such interventions can clarify women’s beliefs, attitudes by educating them on the importance of breast screening; thereby they will make an informed decision and attend the screening appointment (Bekker et al. 1999). The second approach (Figure 17) starts after the actual screening batch. The NBSS creates the results for the last batch and utilises a report template to export the batch list. The user (via GUI) points to the location of the flat file again to segregate the non-attendees. Once segregation is completed, a new SOAP message is generated using the same procedure as before and transmits it to the respective GPs. This again updates the woman’s medical record of real non-attendance, thereby providing another opportunistic type of intervention to the GPs.

From Figure 17, Phase 1 with a two-pronged approach proved to be an ideal choice. Approach 1 is related to the knowledge created through the AI-ATT prediction algorithm that predicts the non-attendance in the pre-screening stage. This approach represents the process of duplicating the knowledge structure personified through the assistance of neural networks. Approach 2 was related to the knowledge from the post-screening results. Both results would be seamlessly integrated with the primary care providers. If the knowledge
from these approaches were to be channelled through an established primary care delivery mechanism (such as a GP surgery), would yield better returns (Richards et al. 2001). Phase 2 is related to the knowledge personification at the primary care delivery end.

**HL7 Message design**

Based on a comprehensive requirements analysis, a conceptual diagram specifying classes and their association with multiplicity for the scenario was constructed as the first step of the message design (Figure 18). Screening units are responsible to screen the eligible female cohorts. Each woman has many screening episodes within the specified age limit.

Women aged more than 70 are automatically out of the recall cycle but can still avail themselves of the screening services by self referral; hence the number of episodes for each woman cannot be limited to a logical number through multiplicity. Each screening unit caters to many GP surgeries and each surgery is serviced by one or more physicians (GPs). Each screening episode has one screening report. This report is mailed to screened women (including non-attendees) and a copy of the report is also sent to the GP surgery with which the particular woman is registered. This class diagram was further modified to reflect the HL7 V3 Reference Information Model (RIM) paradigm.

![Class diagram for screening scenario](image-url)
Modelling methodology

The sample screening report (hard-copy) sent to GPs was the basis on which the message was designed (Figure 20). As the messages developed for the current research specifically addresses the issues pertaining to non-attendance (and being prototype in nature), the focus is on the two proposed scenarios alone (Figure 19).

![Diagram](image)

Figure 19 - Design of message (screening) with objects and relationships

This shortens the scope and avoids the creation of D-MIM (Domain- Message Information Model) As R-MIM is concrete and message specific in nature, the design phase concentrates on creating the R-MIM (Reference Message Information Model) alone for this scenario. An R-MIM pertaining to the pre-screening (prediction) and another related to the post-screening is designed in this phase. The messages developed based on these R-MIMs will be validated and tested through instantiation at the final stage of prototype implementation.

Information analysis

The twin message exchanges planned for this scenario have been developed after extensive discussions and interviews with the participating screening unit staff and the paper-based screening report sent to the GPs has been utilised as the main reference. Based on the details depicted in Figure 19, two “storyboard” cases were created for each of the knowledge transfers (refer to Error! Reference source not found. for more details).
Figure 20 - Sample screening report sent to GP in hard-copy format
The interaction diagram (Figure 21) refers to these storyboard cases for identifying the communication between sender and receiver roles with related triggers. The naming of all messaging components (roles, participants, triggers, interactions and messages etc) follows HL7 naming conventions.

**Figure 21 - Interaction diagram for the twin messages**

**Information modelling**

HL7 V3 message design is model-driven and is achieved through extensive use of graphic models and each model represents a concept in message exchange. In concurrence with this approach and, based on the storyboard, interaction diagram and tables corresponding to RIM classes are cloned. The cloning of classes should reflect the semantic content of the message intended for exchange. This cloning yields the first information model. After meticulous iterations and incremental changes made to the basic model a final diagram representing the complete semantic content of the message exchange can be drawn (Figure 19). The objects reflect the colour scheme as recommended by HL7. The pre-screening message (PRSC_MT060001 – refer to Table 2) is triggered by the trigger event (PRSC_TE070001). The generic class groups related to the basic RIM classes are represented as “swim lanes” for easy understanding and the lines connecting the objects show the relationships between them.

<table>
<thead>
<tr>
<th>Sending Role</th>
<th>Breast Screening Office</th>
<th>PRSC_AR070001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiving Role</td>
<td>General Physician (GP)</td>
<td>PRSC_AR070002</td>
</tr>
<tr>
<td>Trigger Event</td>
<td>Prediction Completed</td>
<td>PRSC_TE070001</td>
</tr>
<tr>
<td>Transmission Wrapper</td>
<td>Send Message Payload</td>
<td>MCCI_RM070001</td>
</tr>
<tr>
<td>Control Act Wrapper</td>
<td>Control Act with author participation</td>
<td>MCAI_RM070001</td>
</tr>
<tr>
<td>Message Type</td>
<td>Pre-Screening Report</td>
<td>PRSC_MT060001</td>
</tr>
</tbody>
</table>
Table 2 - Pre-Screening interaction table (PRSC_IN060001)

The design process identifies the primary act as the “Screening (GP report)”. The roles played within the act are:

- the screening authority (sending role – PRSC_AR070001 – refer to Table 3) sending the message
- the GP as the message recipient (receiving role PRSC_AR070002)
- service delivery location role identifies the place of screening (screening unit or mobile unit)

<table>
<thead>
<tr>
<th>Sending Role</th>
<th>Breast Screening Office</th>
<th>PRSC_AR070001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiving Role</td>
<td>General Physician (GP)</td>
<td>PRSC_AR070002</td>
</tr>
<tr>
<td>Trigger Event</td>
<td>Screening Completed</td>
<td>PRSC_TE070001</td>
</tr>
<tr>
<td>Transmission Wrapper</td>
<td>Send Message Payload</td>
<td>MCCI_RM070001</td>
</tr>
<tr>
<td>Control Act Wrapper</td>
<td>Control Act with author</td>
<td>MCAI_RM070001</td>
</tr>
<tr>
<td>Message Type</td>
<td>Post-Screening Report</td>
<td>PRSC_MT060002</td>
</tr>
</tbody>
</table>

Table 3 - Post-screening interaction table (PRSC_IN060002)

The entities associated with the roles played are individuals represented as persons (A, B and C) and the screening unit. The GP plays the message receiver role. An additional role is assigned to the GP as the family physician for the screened women, which is represented through role relationship “registered with” (Figure 19). The act itself is made up of more than one related act and this pertains to details such as screening type, screening opinion and a note. These acts are related to the screening GP report by act relationship defined as pertinent information for each associated act.

HL7 v3 Message implementation

The message designed in section 4.6 was implemented through the following steps using the specified software tools. The initial message implementation was through UML diagrams (refer to section). With reference to these diagrams, detailed RIM-based domain diagrams were generated. The final design has been translated into an HL7 object model artefact using MS® VISIO® (Professional 2002 v10.0.525) software using HL7 templates (RMIM2) which loads the RIM and vocabulary for the attributes linked to these stencils (version 2.99). The relevant attributes are selected from the RIM classes satisfying the mandatory conditions stipulated by HL7 (such as Instance Identifier, class code, mode code etc) and those relevant to the message content, with reference to the class diagram (Figure 18). Data types, coded elements and multiplicity associated with the relationships were also defined during this phase.

The HL7 UK and Connecting for Health (CfH) have well defined Common Message Element Types (CMETs) within the domain. These CMETs aim at creating ready made blocks for frequently reused components and messaging structures. This facilitates rapid message designing and assists designers in not reinventing-the-wheel. CMETs also provide error-free...
message design with consistency. To gain better understanding of the message designing nuances and to gain more design experience (owing to the projects focussed context) the author created some of the components exclusively for this work, in spite of the existing equivalent CMETs. The wrappers necessary for the messages, Control act wrapper (MCAL_RM070001 refer to Table 1) and Transmission wrapper (MCCL_RM070001 refer to Error! Reference source not found.) were manually created with the VISIO® templates in accordance to CfH MIM (v 3.1.04).

Drawing the control act wrapper as recommended by CfH is simple with the VISIO® stencils, but the transmission wrapper similar to the recommended one has not been possible due to the non-availability of certain library files. This created validation errors in VISIO®; failing this validation, no Hierarchical Message Definitions (HMDs) were created. Hence, the automated schema generator (HL7 V3 Schema Builder V1.20) is not able to generate schemas. To overcome this problem, and to complete the message design for the prototype, the wrapper schema has been hand crafted. The rest of the message components are validated using the HL7 vocabulary and HMDs are built using “RoseTree” software (version 2.9.55, RIM-V02-04). Schema Builder has been used to generate the XML schemas automatically (refer to Error! Reference source not found. and Error! Reference source not found. for a sample schema for the proposed messages). Similar to the pre-screening message and its components, the trigger event, message type and its associated sending and receiving roles have been designed for post-screening interaction (PRSC_IN060002, refer to Table 3). The same control act wrapper and transmission wrapper have been reused for this message.

**Challenges faced during message design and implementation phase**

HL7 standard version 3 is a complex standard for novice users. Exposure to both healthcare and software domains would greatly assist in understanding the more rudimentary aspects of HL7 version 3. This project had proved that even a novice in EDI and HL7 could make substantial inroads if the learning process were to be properly directed.

The following critical aspects were encountered during the research, this type of HL7 V3 message design process is recommended where the available resources are limited.

1. Becoming a member in the local HL7 organisation has its own benefit. It is a good place to meet and know the local thrust areas and the modalities for local implementation. An important factor would be to get some form of introduction to the standards through professionally designed training sessions which are organised by the HL7 group.

2. There is a sizeable volume of literature available at various sites, such as HL7 UK and HL7 (US, Canada and Australia to name a few).

3. The HL7 parent organisation offers tools and technologies for developers and message designers and this provides a conducive environment for beginners to get a foot hold easily in design and implementation of HL7 messages.
4. Not only the HL7 V3 development tools are provided free for download even CfH MIM (for members only) is also available for new message developers. These support mechanisms also include necessary guidelines including configuration and installation notes.

During the initial phase of developing the HL7 version 3, the Standards Development Organisation (SDO) for HL7 had pursued an easier method for setting up the working environment, hence had logically selected a well established, off-the-shelf tool such as MS® VISIO® and added its own libraries through Visual Basic (VB) based macros. Rose Tree offers the much needed interface to convert the VISIO® designs into HMDs and message instances. These are exported as comma separated values (CSV) for further processing. Schema Builder, a java based programme takes in these CSV files as input to generate XML schemas automatically. There has been a migration towards the use of Eclipse as a development environment and this has gained popularity but this is not mandatory for message development.

Identifying the related HL7 library files and having access to them can be a daunting challenge for any outsiders to the working group. Hence some form of collaboration with the local (country level) sub-group is vital for any message development. This would avoid a lot of hardship and align the projects with the local implementation. Even though the VISIO® stencils, Rose Tree and Schema Builder are free to download, the need for MS® VISIO® which is commercially available software is necessary for designing the messages and this can be a handicap as it is not free to download. The literature and other related papers are spread across different domains and thus require careful collation for effective comprehension. The lack of exclusive academic courses on HL7 or healthcare informatics has created a gap in training and educating fresh talents at graduate level in this field. In spite of the freely available tools, there are a few barriers to be overcome as a beginner. These tools were developed by individual groups within the HL7 SDO and hence function as separate components. An enterprise-based solution which encompasses all the tools and a single interface would be very useful. This would reduce the complexities of the message design and implementation.

This particular message design itself was simple and closely corresponded to the paper based communication between the screening unit and GPs. Mapping the contents and adding new components to satisfy the HL7 syntax without affecting the semantics was relatively easy. For novice designers, such message design can be a good starting point. As the designers gain more experience, more complicated messages (for example, spanning multiple clinical documents) within the domain can be designed with relative ease. The scope of this project had a distinct focus and therefore provided a unique opportunity to complete the whole messaging (design through implementation). This research is directed at creating a prototype level message only. The final prototype would be a complete software environment which would incorporate a database input hooked to the AI network for predicting non-attendance. The predicted women will be automatically associated with the respective GPs and the message content would be transferred to the messaging module. This module will use the message content and the schema (as designed and developed earlier) to generate XML instances. These instances would be part of a Simple Object Access Protocol (SOAP) message and sent to the respective GP system. This would later be utilised to initiate GP intervention when the identified women approach the GPs for other health-related services.

All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org
This section enumerated some of the challenges that can be encountered while creating HL7 messages in a highly focussed environment such as the prototype discussed. The author has exhibited that HL7 development methodology could be learned and applied in a short time. As the standards are maturing and more countries are adopting and implementing HL7 V3, this research can be an ideal starting point for novice message designers and implementers.

**Proposed protocol**

Reasons for non-attendance may be largely attributed to disinterest in attending a mammography session, negative attitudes, beliefs, prior or current medical problems, and fear of X-rays (Bekker et al. 1999). These reasons can be negated by a proper education provided to women. Education has to be directed at explaining the advantages and importance of screening and assist in removing the socio-cultural and personal barriers (Cassandra 2006). Other possible options including convenience in terms of time, place and dates provided to women for encouraging their attendance. In spite of the expedient measures provided to the women, non-attendance has been a grave concern for the National Screening Programme. This scenario can be properly addressed if women can be identified who may probably not attend in advance and directing additional resources to educate these women, thereby increasing attendance. To summarise the messaging and the practicality of the proposed solution, a detailed architecture utilising the existing technologies and proposed components are recapitulated in this section as a new breast screening protocol. The same architecture can be mapped for other message paths in the future.

**Proposed Breast Screening Attendance Messaging Protocol (BSAMP)**

The Health Level Seven (HL7) Standard is the *de facto* standard for all electronic exchange of data with in healthcare environments in the NHS (Thorp 2005). The HL7 standard defines seamless exchange of actual data (and its structure) between health care applications (Beeler 1998). This section addresses the BSAMP and its messages pertaining to a prototype model created with the assistance of an Artificial Intelligence (AI) program (AI layer - refer to Figure 22) to predict breast screening non-attendees and to share this knowledge with the primary care deliverers (namely the GPs) and they are:

- Pre-screening (predicted non-attendance)
- Post-screening (actual non-attendees)

These two approaches were described in the earlier sections with clarity. The driving component for both the approaches is the infrastructure required for the knowledge sharing. The existing technologies described have to be utilised to reflect the IT revolution occurring under the NHS’ national initiative. This would not only provide the path for faster implementation but also reduce the required overheads by utilising the in-place system for knowledge exchange. The proposed BSAMP protocol leverages these technologies as shown in Figure 22. This illustration also points at five different message flow paths associated with the said screening domain.
‘Path 1’ is the existing messaging channel used by the current screening centre’s exchange of screening report (at post-screening stage) to the GPs in hard-copy format. This path is cumbersome not only to produce at the screening centre’s administration office but also would require considerable manpower and resources at the GP end for entering the information in to the women’s clinical records (either digital or paper-based). The remaining paths are all proposed to utilise the new IT infrastructure for other message exchanges in concurrence with the suggested protocol. The screening centre collates all its screening statistics in to consolidated information package for ‘Körner Community’ returns (KC62 and KC63 reports). These reports are sent to the National Screening Programmes for generating national statistics for all screening programmes. This information exchange can be made using the same architecture as portrayed in Figure 22, ‘path 2’.

‘Path 3’ proposes to download and upload any changes to the women’s demography and personnel information. ‘Path 4’ is intended to record the outcomes of the screening in to the Personnel Spine Information Service (PSIS). This updating will provide an opportunity for all the healthcare providers to gain knowledge of the screening outcome (including attendance and non-attendance) so that a collective effort can be made to deliver better health care. ‘Path 5’ is for sharing the knowledge of non-attendance gathered from pre-screening (prediction) and post-screening report. The NBSS system communicates the dataset to the AI layer for prediction, and the outcome is appropriately packaged as HL7-based SOAP messages by the wrapper layer and then via the Data Transfer Service (DTS) client interface, the message is sent to the DTS layer. This populates the respective GP mail boxes.
Typically, on regular intervals (3 to 5 times per day) the GP application layer executes a batch process of uploading and downloading the GP mail box through the DTS client interface for GPs. When the screening report message arrives at the GP application layer, after due authentication, the individual women’s clinical records are updated. There are two main families of GP software for assisting GPs in their clinical consultations. One type of GP software uses pop-up windows and another uses framed structures to draw attention through textual prompts (Figure 23 and Figure 24).

![Figure 23 - Screen shot of prediction and non-attendance displayed in frames with different set of font colours](image)

When the woman approaches the GP for consultation, the GP retrieves the women’s record and these trigger the automated prompts (Figure 24). The prompts are amber coloured and pop up to remind the GP to intervene, since this woman has been predicted as a non-attendee. The other screen shot (Figure 24) depicts the red coloured pop-up prompt signifying the non-attendance at post-screening stage. This incidentally adds value to the already existing prompt (amber coloured) of predicted non-attendance. In the other type of frame-based GUI, appropriate coloured text is used to highlight the importance of the shared knowledge.
Conclusion

The BSAMP architecture relies on existing technologies. This gains significance as it implies that this is an implementable solution. The results collected through the qualitative and quantitative analysis served to further understanding as to how knowledge creation and sharing can be established with reasonable efficiency. The AI triangulation provided the efficiency metrics (benchmarks) for future improvements. The NCRS is already being tested across the country. The architecture calls for close cooperation with the screening units and other care deliverers. Such cooperative efforts can make significant impact on screening attendance. This chapter describes how KM-based tools and techniques can assist in evolving new healthcare strategies to address new challenges.
Algorithm for Screening Attendance Prediction

The first step towards addressing the screening attendance was largely completed by Arochena’s (2003) research on predicting non-attending women using an Artificial Intelligence (AI) based predicting algorithm. The following section provides a summary of the Arochena’s work (AI-ATT), a brief analysis of its functioning and working environment, later the chapter focuses on the proposed (JAABS) algorithm’s data pre-processing modules and finally dwells on the JAABS algorithm’s design aspects. As discussed earlier, quantitative analysis was completed on the AI’s performance metrics. Further, these results were triangulated to reflect how the new algorithm compares with the earlier one and also to the statistical technique (logistic regression). This comparison of the JAABS vs. AI-ATT is enumerated and concluded.

Predecessor algorithm employed in exploratory research

The research by Arochena (2003), funded by and tested in the Warwickshire, Solihull and Coventry Breast Screening Unit, (project ID 2000/112) indicated that women’s attendance can be predicted by a software system using the Artificial Intelligence based attendance predicting algorithm (AI-ATT). The algorithm has been evaluated with individual AI methods (feed forward and radial basis function), and statistical methods (such as Logistic Regression - LR) and was found to be extremely favourable in predicting attendance (Arochena 2003). This algorithm can assist the Breast Screening Unit (BSU) to effectively schedule the screening programme and eventually result in the efficient use of limited available resources. Even if the attendance percentage were increased marginally, a significant number of lives may be saved. Arochena’s (2003) research resulted in the formulation of the required artificial intelligence-based attendance prediction algorithm and was trained and tested on datasets collected during a ten-year period at the Warwickshire, Solihull and Coventry Breast Screening Unit. The blind study (data not seen earlier by the AI) test results were encouraging, especially from the second episode onwards, warranting the algorithm to possess enough potential to be incorporated in a Breast Screening Programme (BSP) framework.

Wilson’s (1999) work focussed in developing a methodology to extract data from the MUMPS database. The same methodology was later used to extract dataset pertaining to the period of 1989 to 2001 from the Warwickshire, Solihull and Coventry Breast Screening Unit for testing and validating the AI-ATT. In addition, the National Screening Programme has been constantly striving to provide better services to the public and one of the new enhancements offered by the screening services is to increase the screening age limit from sixty four to seventy (Patnick 2006). This effectively increased the number of screening episodes and resulted in augmenting the need for effective use of the already stretched NHS resources.
All the above said factors raise the important question pertaining to how effectively the predicted knowledge can be leveraged and managed.

An exploratory research exercise identified two algorithms, one related to predicting attendance on screening batch and the other to predict the incidence of screening variation, i.e., changes to the appointment dates, resulting in late attendance (Arochena et al. 2001). The research further validates the algorithm through a blind study by using the latest data from the screening unit (not formerly seen by the algorithm while training and testing) (Arochena et al. 2003, Aochena et al. 2002). The results confirmed that these algorithms have significant potential and were predicting with reasonable accuracy. This warranted further research to confirm the feasibility of implementing such algorithms in the routine HC delivery process at the screening unit.

**Artificial Intelligence based Attendance Prediction algorithm (AI-ATT)**

The earlier section gave a brief overview on the research completed by Arochena (2003). This work had modelled the predicting factors and was validated through an Artificial Intelligence (AI) based predicting algorithm. The dataset (containing 281,415 potential screenings related to 147,432 women) was downloaded from the screening unit’s MUMPS (Massachusetts General Hospital Utility Multi-Programming System) database (Wilson 1999, Arochena et al. 2002). This data was filtered by a combination of Visual Basic® (VB) and Sequential Query Language (SQL) based macros and routines in a Microsoft Excel environment respectively. The routines used in this data filtering process were primarily designed to identify unknown date of birth, irregularities in date entry, incomplete screening end code, inconsistent episode numbers, anomalous SX number, and other missing data, incorrect data entry, duplicate entries etc (Wilson 1999).

**Analysis of AI-ATT**

Considerable association was acknowledged between age bands, attendance and screening end code (Arochena et al. 2003, Arochena et al. 2003). Age band including round length (difference in years between the date of first offered appointment of one screening episode and the following one), previous screening attendance and screening variation had high association.
Postal area, Townsend deprivation, screening variation, past cancer history, false positive, history of false positive and type of invitation (first call or recall) to the episode also had substantial relationship with the screening attendance (Arochena 2003, Arochena et al. 2003). The predicting factors were not the same for all the episodes due to various reasons. This prompted the input data to be split into their respective episode groups. The algorithm has been evaluated and compared with Logistic Regression (LR) model and was found to be more precise in predicting women attendance (Arochena 2003, Arochena et al. 2001). The algorithm employs two approaches one through a neural network pruning and the other by radial basis function based network. A hybrid of these two approaches was employed in the AI-ATT to attain better performance metrics like accuracy, sensitivity, specificity and positive & negative predicted values when compared to LR (Arochena 2003, Arochena et al. 2001).

**AI-ATT working environment**

The analyses of relationships between the input variables and the attendance pointed that not all the episodes had the same set of predictor variables. For example the first episode which has all the relevant demographic details of the women lacked screening history which was a handicap and yielded less accurate results (Arochena 2003). After exhaustive analyses, it was found that the type of invitation and the age of the women were positive prediction attributes for the first episode group (Arochena 2003). Whereas the remaining episodes comprised of more details (demographic and screening history) this enabled better attendance prediction. The attribute indicating non-attendance in previous episode and diagnoses of breast cancer in the previous screening episode were also identified as
predicting factors (Arochena et al. 2002). The algorithm functions are enumerated as steps one through seven (Figure 26). The dataset as shown is derived from two different sources, the local Primary Care Trust (PCT) holds the updated information of the women's demographic details and the screening unit holds the details pertaining to the women's past screening episode details. The screening unit downloads the demographic data from the PCT to create the batch list for the upcoming episode. The algorithm then combines this data with the screening history for available women. The data from this combination is extracted as a single file which is in Microsoft compatible format for further processing. The data is manipulated and reorganised based on the episode number and each woman is identified with an episode number which is inserted as a variable.

![Figure 26 - AI-ATT algorithm working process (Arochena 2003)](image-url)
Based on this inserted episode number the single monolithic file is partitioned into respective episode groups and saved as sub-files (Arochena 2003). The postcode of the women’s residence (extracted from the women’s demographic data) at the screening time period is matched with the “Townsend’s” deprivation score and is stored as a new variable.

Townsend’s deprivation score is calculated utilising the census data. The time difference between the census data and the physical expansion rate of a city results in areas that do not have corresponding deprivation scores. It was found that not all the postcodes had an associated Townsend’s deprivation score; hence based on the presence of this variable the sub-file is further classified (Arochena et al. 2002). A number derived from the postal area enumeration based on district list is assigned for each record. This number is leveraged for postcodes that do not have related Townsend’s deprivation score and is stored as a new variable called post annum. The above said two sub-files (one related to Townsend’s deprivation score and another related to postcode) are individually fed to two Artificial Neural Network models (ANN) for every episode. ANN based on Pruning (ANNP) and Radial Basis Function Network (RBFN) are the two ANNs employed. These ANNs generate two individual variables; one related to the attendance prediction and another related to the confidence value of the prediction (Arochena 2003).

The final activity is related to the classification of the outcome from the previous step. The confidence level of the predicted attendance is compared between the values generated by the ANNP and RBFN algorithm. The highest confidence level between the two is voted as the final prediction of the AI-ATT algorithm. The final dataset is again collated from these sub-files into one consolidated output file with the predicted attendance (Arochena 2003, Arochena et al. 2000, 2001, 2002, 2003). The dataset collected during the ten-year period had many versions or styles which need to be regularised towards a common platform before being put to use within the algorithm (Arochena et al. 2000). This data processing was carried out using several different software data analysis tools. These tools included Microsoft Excel®, Visual Basic® routines and Statistical Package for the Social Sciences (SPSS®) and Clementine® to extract the data and adapt to the AI models. The algorithm itself was modelled in a visual modelling environment called Clementine® (Arochena 2003).

The algorithm for its proper functioning depends on all the aforementioned software. This posed a challenge in implementing the algorithm within the routine process of the breast screening office environment. The primary challenge was that they do not possess the required licenses for the said software nor the expertise in linking all the software to work seamlessly for predicting on a monthly basis for every screening batch. The earlier research by Arochena (2003) was limited to identifying whether the non-attendees can be predicted with reliable accuracy and was not aimed at the various facets of implementing the algorithm nor on envisaging how to effectively use the gained knowledge of probable non-attendance. This project work addresses such issues from an implementation point view and suggests how this knowledge can be leveraged. The qualitative research conducted at the initial phase through interviews fostered better understanding of the previous (AI-ATT) algorithm’s developmental challenges and how they were resolved. This played a significant role in reducing the development time for the new algorithm. The next section enumerates the design and implementation of the new algorithm.
Data pre-processing

The introductory section gave a brief insight to the research by Arochen (2003), who had pursued an exploratory type of research focused on identifying the attributes that can be classified as predicting factors. The primary objective of the said research was to propose the predicting algorithm which includes the modelling of the prediction features that affect screening attendance to screening invitations. The current work elaborates the need for an open source-based data pre-processing module (Figure 27). It further illustrates the design of this module and the variables participating in the prediction process.

Data Pre-processing design

The demographic details for the three year call/recall are downloaded from the local health care authority’s database. The downloading is effected via the health link network on to a standalone system within the Breast Screening Unit (BSU). The historical data related to screening, appointments and results pertaining to screening women are retained within the screening unit’s “MUMPS” database (Figure 27). “MUMPS” also called the Oxford system, is one of the earliest programming languages used since 1960’s (ÍOâne 2002). This language was extensively employed to write database applications explicitly for the HC domain.

The MUMPS database works on Disk Operating System (DOS) and employs a character-based user interface for database interrogation (ÓKane 2002). The cumbersome DOS based system was prone to erroneous data entry and hence warranted a change in the system. New software, National Breast Screening Computer System (NBSS), was developed in 2002-3 to address these issues (Baskaran at al. 2005, Tarver et al. 2004).

This NBSS consists of a Visual Basic® (VB) front end connected to a “Caché® database which was seamlessly integrated with the MUMPS database (Tarver et al. 2004). All the
above processes required multiple software interactions, functioning on different environments which resulted in considerable complexities during data extraction for the current project.

**Module design**

The earlier research conducted within the BIOCORE had used the MUMPS database to extract data. This data was saved as a flat file and then imported into MS EXCEL® for further processing. The data was filtered by VB® application routines and SQL queries which were manually executed and verified for its correctness (Arochena 2003, Wilson 1999). The filtering addressed Y2K (year 2000) bug, unknown date of birth, irregular data entries, incomplete screening end code, anomalous episode number etc (Arochena 2003). The above said process was tedious and human centric. Each and every record has to be validated manually and incomplete dataset were excluded from the study. In the current study very similar difficulties were encountered. In addition to these challenges, the prototype proposed has to be completely automated. The automation has to address the removal of anomalous data, wrong data entry, duplicate entries etc. All the experts’ involvement in data validation and analyses has to be expedited through the prototype with little or no human intervention.

In an ideal scenario, the functioning of the proposed prototype (Figure 27) for attendance prediction would encompass all the above data processing activities plus the generation of the required predictor variables using the logic proposed in the algorithm. The data pre-processing module after completing the above said activities has to generate a new dataset that would be ready to be assimilated by the AI module. The earlier research’s data processing was conducted in different environments hence was difficult to be integrated as one whole functioning unit. All the above said challenges can be solved if there is one software module which can deliver a complete and automated data pre-processing for the prototype. Hence an automated data pre-processing is indispensable for the successful functioning of the proposed algorithm. All the activities pertaining to the proposed algorithm can be summarised as one ‘use case’ diagram (Figure 28).
Variables identified for data pre-processing

The data pre-processing module can be visualised as two separate sub-components. The screening office module is executed with the existing software programmes available in the breast screening office and the new Java-based pre-processor module for collating the data. The VB® front end made data extraction straightforward (Figure 27) from the MUMPS database through SQL queries directed at the Caché® database. Currently the breast screening office is employing “Crystal Report” (CR) as part of the NBSS to generate reports for all the screening activities including screening, administration, invitation etc. Part of the data pre-processing was implemented through CR software. The screening unit had earlier indicated that the routine functioning of the screening office should not be affected during the data extraction process. Hence, prior to the data extraction a CR template was created to reflect the format of the data to be exported. This template was used to export the data as a flat file. This strategy was adopted since all the screening units across the country were expected to have some form of minimum facility for creating dataset in a flat file format.

Coupled with this a need for a low overhead on the existing IT system and minimum additional complexities were expected as fundamental for the proposed prototype. All the aforementioned rationale strengthened the need for adopting a compromise strategy which would export data as a flat file, so that the mode of data transfer can be standardised across the country with minimum or no interrogation with the screening database. The CR template generates the dataset as depicted in the format (Table 4). The query generated the demographic and episode details for all the women in as many records. The demographic data was not complete and only the first record of a particular woman had the complete dataset and the remaining records of the women corresponded to the historical episode details. The SX number was repeated one for the demographic data and one for each episode.
The SX number is the link between the two components of the record. The women’s address and name are excluded from the study to address data protection and to maintain anonymity. In spite of the necessity of the messaging module the complete data set was generated without the personal information of the screening women. The post code of the women is indispensable for the current study, as it generates the important predictor variable in the form of Townsend’s reference number and post annum number. The recipient of the proposed knowledge transfer is the woman’s GP, hence GP information like surname, surgery address and postcode are necessary for sending the HL7 based EDI message. This dataset including the women’s GP details are collated as a separate entity. For further use, it is stored as a look-up file for generating messages. The SX number acts as the link between these two datasets. The women’s demography is represented as a ‘Record’ object (Figure 29). The screening detail for each episode is represented as ‘Episode’ object.

One ‘Record’ object would be associated with one or more ‘Episode’ objects. The gaps in the demographic record have to be filled and the episode details are associated with the women’s demographic data. Exhaustive analyses of the data indicated that the CR report had duplicate episode details. Upon further analysis it was found that rescheduling the appointment by the women was the primary reason for such multiple and duplicate entries. These entries are to be removed before further processing can be implemented. Each record
read from the CR report has to be first partitioned in to episode details and are stored as ‘Episode’ objects. They are finally collated and associated with the women’s demographic details (represented as ‘Record’ object). In addition to this all the records has to be automatically validated. The earlier work by Arochena (2003) had identified all the contributing predictor attributes through comprehensive statistical analyses. Based on these analyses, the following attributes are generated in this module.

- Townsend reference number
- Post annum number
- Age band (age of the women for the current episode, converted as categorical variable)
- Previous screening attendance
- Slip (difference between the date of first offered appointment and the actual screening)
- Number of tests (diagnoses corresponding to the immediately prior episode)
- Cancer detection (in the immediately prior episode)
- False positive (false positive diagnoses results in the immediately prior episode)
- History of false positive (false positive occurrence in any of the previous episodes)
- History of cancer (cancer detection in any of the previous episodes)

After generating these attributes the pre-processor module classifies the ‘Record’ objects based on the number of ‘Episode’ objects it contains. This dataset is then written as an in-process flat file for reference. All errors generated during the execution of the pre-processing module are written in a log (error) and is also saved as a flat file for future reference.

**Data pre-processing module implementation**

The data pre-processing module is built with a sub-module called “IO module” (Figure 27) for reusability. This module has three classes, one to create a file for a given name, another to create a stream to read the file line-by-line and the last being the class to invoke a file writing stream. The main class of the data pre-processing module is the “DataPreProcessor” class. This class drives the whole module; a ‘Record’ class stores the attributes of the screening women’s demographic details. The ‘Episode’ class holds the attributes of the woman’s screening history details. The ‘Record’ class stores all the episodes it is related to, in a “Vector” object. The ‘Record’ class uses “DateCompare” class for comparing dates to store the episodes sequentially and to identify ‘Episode’ object duplicates. The same class is used to generate the “Age band” predictor variable from the woman’s age and the date of first offered appointment and “Slip” variable is also calculated using the same class by finding the difference of date between the appointment date and the actual screening date in days.

The “DataPreProcessor” class populates vector objects with the Townsend reference and “Postannum” number by reading values form a flat file utilising the “IO Module”. The
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

postcode’s first part is extracted and used to check the availability of Townsend number stored as Vector object. If the Townsend number is not available then an arbitrary number (say 100) is stored as Townsend reference number. Similar to this a “Postannum” number derived from the postal area enumeration is matched with the district list (from a lookup file) and stored in the ‘Record’ object. The “DataPreProcessor” object first loads the input flat file; the flat file could be either as comma separated values (CSV) or tab delimited values. To load this file the class utilises the IO module’s “ProcessFileReader” class and reads line by line and process one line at a time to create the ‘Record’ and ‘Episode’ objects respectively until the end of file. After creating these objects, the predictor attributes are also generated based on the input data and stores in their respective ‘Episode’ objects. The “DataPreProcessor” object then saves the data by opening a file writer stream through the “ProcessFileWriter” object. This file becomes the input to the subsequent AI module for normalisation of data before feeding it to the neural network.

Proposed algorithm

During the initial stages, qualitative analysis revealed that the software platform was crucial for the success of the prototype. The need for developing the prototype in an open source environment was also recognised. The development of the prototype in open source technology would assist in easy integration of the algorithm within the breast screening services. This strategy would not force the screening services on any additional financial overheads for adopting this algorithm in the form of license fees or new software tools.

Java based Attendance prediction by Artificial Intelligence for Breast Screening (JAABS) algorithm

“JAABS” is the new algorithm designed and developed in JAVA ® based technology. All the screening services has to have is a Java Runtime Environment (JRE) (which is free to download) for running the algorithm (Griswold and Lawrence 2004).

As the design process was based on more of an evolutionary type, a modular design strategy was selected. This would assist in parallel development of the implementation and also enable testing as modules rather than one single monolithic type of programme. The modular design also ensured that any additions or changes can be implemented without affecting the other modules. This was possible by the Java’s Object Oriented technology (Griswold and Lawrence 2004).

JAABS design

The “Artificial Intelligence Module” encompasses the data normaliser; the neural networks and the results collator (Figure 30). As its predecessor (AI-ATT), the Java based algorithm implements two different neural networks, feed forward Back-propagation Prunning Neural Network (BPNN) and Radial Basis Function Neural network (RBFN). The neural networks proposed needs the input data vector classified as binary values. This conversion of data
into binary is termed as normalisation. Each possible outcome in the input data is represented as a binary value, a “1” indicates the presence of the value and alternatively “0” represents the absence.

The basic neural network and its back propagation-based learning technique are common for the two networks (BPNN & RBFN). The input data in the RBFN is first passed through a radial basis function algorithm, to identify the clusters and assign a radius for cluster classification. These cluster centres are calculated and the real time data is checked against these established clusters centres. The distance of the data point vector or record within the input data space is arbitrarily assigned to the nearest cluster group. This data is then fed to the neural network for making the prediction of attendance. Each episode has a different set of predictor attributes; hence each episode is fed through separate neural networks which were trained with their respective training dataset (Negnevitsky 2004, Haykin 1999, Wasserman 1993, MacClelland and Rumelhart 1986, Bigus and Bigus 2001, Anderson 1995, Masters 1995, Hudson and Cohen 2000, Moor 2003, Joos 2005).

Each dataset classified earlier into those with Townsend reference number and without (this data set has the “Postannum” as the predictor variable) is used as the input for the normaliser. This normalised data set is fed into the stream to generate the predictions. All the results, one each from the Townsend reference and “Postannum” number stream of output from the BPNN and RBFN network are collated by the “Results collator”. The results module collects the collated prediction for each episode and submits to a “Poller” based classifier. The “Poller” finds the best prediction for the given episode, for each stream of dataset with Townsend reference number or without. The “Poller” generates the final prediction output based on the confidence value of the prediction. This is fed in to the prediction results collator for all the input (women) based on each episode. This is associated with the women SX number and NHS number. The consolidated result is used to generate the non-attendance list and written as a flat file for processing by the “Messaging Module” for message generation.
JAABS classes

The JAABS algorithm is designed as shown in Figure 31. The interface for the whole algorithm is through “JAABSAlgorithmInterface” class. This class provides the sequential logic for implementing the feed forward back propagation pruning neural network and the radial basis function neural network. This module also utilises the IO module (sub module) of the Data Pre-processing Module for input and output functions on dataset. The “JAABSAlgorithmInterface” class instantiates two normalising classes, one each for BPNN and RBFN respectively. The normalisation objects accept input file name and output file name as parameters. This generates the collated predictor attributes with binary values for feeding into the neural networks.

![Figure 31 - JAABS algorithm (class diagram)](image)

The BPNN neural network is implemented through the “BPNNAlgorithm” class. This class works on a multilayer feed forward model. This class is instantiated by passing the following parameters; number of input nodes, hidden nodes, output nodes and values for learning rate, momentum and tolerance for training the network respectively. The multilayer feed forward network refers to the predictor variables passing through the input nodes, hidden nodes and to the output node as a stream. During the training mode, a convergence strategy is adopted. In this strategy, a small proportion of the error value is back propagated until the network error is within specified limits or if it completes the
number of specified epochs (training cycles). After the training is completed, the network is exposed to the validation data set (refer to later sections for details of training, validation, and test dataset generation) and the predictions are saved as a flat file. The same way the final testing set (data not exposed during the training and validating sequence) is fed through the network to get the prediction of previously unseen dataset by the network.

The JAABS interface sequentially instantiates the RBFN stream normalisation and feeds the binary data set to the “RBFNAlogrithm” class as “DataPoint” objects. This class instantiates the input data as data vectors which is associated with a cluster through “Cluster” object. A “Radius” object is created to refer to the distance of the input data vector to the cluster centre. A “Centroid” object refers to the centres of each cluster. The complete activities and their interrelationships of the JAABS algorithm stream are shown in Figure 32. The additional episodes and the extensibility to any number of maximum episodes are clearly indicated. The entire diagram provides an understanding of the required degree of automation that has been incorporated within the JAABS algorithm.

Testing JAABS algorithm with simulated dataset

The initial testing of the JAABS algorithm was not tested with real-time data. A simulated data set was created which contained values that can be mapped to the input domain. This testing is aimed only at verifying the functioning of the JAABS algorithm and is not intended to check the efficiency of the algorithm. This testing generated attributes based on the Receiver Operating Characteristics (ROC) analysis (Lavra et al. 2003).
The ROC analysis yielded the figures quoted in Table 5 for True Positive Rate (TPR), False Negative Rate (FNR), True Negative Rate (TNR), False Positive Rate (FPR), Positive Predictive Value (PPV) and Negative Predictive Value (NPV) (Figure 33). Based on the above said metrics further classifier characteristics can be evaluated, like Macro Average (MA), Break Even (BE) and F-Measure (FM) (Lavra et al. 2003).

<table>
<thead>
<tr>
<th>ROC Metrics</th>
<th>BPNN Network</th>
<th>RBFN Network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Validation set</td>
</tr>
<tr>
<td>TPR</td>
<td>100.00%</td>
<td>98.25%</td>
</tr>
<tr>
<td>FNR</td>
<td>0.00%</td>
<td>1.75%</td>
</tr>
<tr>
<td>TNR</td>
<td>100.00%</td>
<td>83.33%</td>
</tr>
<tr>
<td>FPR</td>
<td>0.00%</td>
<td>16.67%</td>
</tr>
<tr>
<td>PPV</td>
<td>100.00%</td>
<td>80.00%</td>
</tr>
<tr>
<td>NPV</td>
<td>100.00%</td>
<td>98.59%</td>
</tr>
<tr>
<td>MV</td>
<td>100.00%</td>
<td>90.79%</td>
</tr>
<tr>
<td>BE</td>
<td>100.00%</td>
<td>89.12%</td>
</tr>
<tr>
<td>FM</td>
<td>0.00%</td>
<td>1.94%</td>
</tr>
</tbody>
</table>

Table 5 - JAABS Algorithm's ROC metrics

Figure 33 - The ROC characteristics for the JAABS algorithm
The Data Pre-processing module processed the flat file and fed the data to the JAABS algorithm (Figure 32). The data within the algorithm was first normalised and fed to the BPNN and RBFN network respectively and prediction from these networks were written in to the output file. The data from the prediction has been analysed in MS EXCEL® to generate the ROC characteristics. The TPR, TNR, PPV, NPV, MV and BE metrics for the BPNN was the maximum for the training set since the network has already been exposed to this dataset. Whereas, the metrics for RBFN training set was not 100% as in the BPNN training set metrics. For the current project, the most important metric in the ROC analyses is the TNR. This metric corresponds to the prediction rate of the non-attendance. The BPNN network TNR rate was only 4% (1) for the test set (Figure 34). The RBFN reported a better TNR for the same test set at forty one percent (41%).

These figures are nowhere near a perfect prediction metric; the reason can be mainly attributed to the fact that the input dataset was a simulated one and the majority of the cases being positive (women that attend), therefore the algorithm learns those cases that have more probability to be true and if not tuned properly, almost does not identify the other cases (non-attendance) at all. The dataset may not represent the a real-time domain values so that prediction can be achieved with reasonable accuracy. Where as the TPR values (Figure 34) are very high, these values are related to the positive prediction of the women who are probable attendees. The RBFN network has a relatively higher TPR rate. BPNN’s TPR is not far behind. This demonstrates that the planned classifier to identify the relatively better prediction outcome among the two networks based on the confidence value can be achieved than any of the algorithms used independently. Also this is in concurrence with the strategy adopted in the predecessor’s (AI-ATT) work for employing such a classifier. Further analysis with a real time data set is described in the later sections of the current chapter.
Comparison of AI-ATT with JAABS implementation

One of the current project’s deliverables is the prototype implementation through the JAABS algorithm. It is imperative that this algorithm be compared with the previous AI-ATT algorithm to identify performance changes (perhaps due to improvements of the JAABS algorithm). This strategy was adopted based on the information collected during the qualitative analysis (open ended questions to the experts and semi structured interviews). The AI-ATT algorithm was designed and validated on the dataset downloaded from the Warwickshire, Solihull and Coventry Breast Screening Unit (Arochena 2003). The dataset was more than five years old and lot of transformations had occurred within the same breast screening unit. The upper age limit for routine screening has been increased to seventy (Patnick 2006). This has resulted in more episodes than the AI-ATT algorithm has considered. The maximum number of episodes used for attendance prediction by AI-ATT was four. The current research has to take into account the possibility of an increased number of episodes, the normal maximum number of episodes for JAABS algorithm has been expected to be seven. The JAABS algorithm can easily be extended to cater to any number of episodes for a given female population. During the qualitative research stage, the breast screening unit’s manager had made a suggestion to identify more predictors to improve the efficacy of the new algorithm. One major contribution to the prediction algorithm made by JAABS in this project work is to identify more attributes. There were several restrictions recognized in this initiative:

- The new attribute should be deduced from the screening unit’s database
- Data security, data integrity and privacy issues should not be breached
- If any processing is mandatory for the new variable, then it should be viable with existing computing technologies
- There should not be put undue resource constraints on the screening unit
- The coding should tie in to the JAABS algorithm
- The independency (non-correlation) hypothesis should be satisfied with the already selected predictors for accurate use of NN

These restrictions made the selection of new predictor variable very complex. Further investigation on the dataset, revealed a possible new attribute in the form of distance between the women’s residence and screening unit’s location for taking the mammogram. The qualitative research established that the postcode of the women’s residence and the postcode of the screening unit can be used to calculate the required distance variable. This distance factor was identified as a predictor after due exploratory tests and later included as one of the predictor attributes in the JAABS algorithm. The validation of this new attribute and the relative performance increase by the new attribute to the screening prediction is enumerated in the later sections. During the qualitative research stage, the breast screening unit’s manager had suggested the identification of more predictors to improve the efficacy of the new algorithm. Further investigation on the dataset, revealed a possible new attribute in the form of distance between the women’s residence and screening unit’s location for taking the mammogram.
The qualitative research established that the postcode of the women’s residence and the postcode of the screening unit can be used to calculate the required distance variable. In addition to this it was also found that the screening unit was at different locations and the mobile units gave the portability factor to the screening equipment. The clinic code was identified as a possible constituent for processing the new distance-based variable. A novel approach of using area code (first part of the postcode) related to the residence and the screening unit (imaging location) was proposed. The “ClinicCode” represented the location of the mobile or stationary units. The screening unit’s postcode (area code) and the women’s postcode (area code) were separated and provided as parameters to the “DistanceCalculator” class to calculate the distance.

The “DistanceCalculator” class matched the area codes to their respective latitudes and longitudes from a lookup file and then the distance was calculated in miles/kilometres. This distance was later categorised as a new variable and named as “Scr dist”. This new distance-based variable was not part of the AI-ATT algorithm and was recommended as one of the future predictor attributes.

JAABS evaluation

The JAABS algorithm was looked upon as an opportunity to maintain the continuity of the earlier exploratory research and for validating the same. Hence JAABS was tested with the data collected from the Warwickshire, Solihull and Coventry Breast Screening Unit, which was also the data source for validating the AI-ATT algorithm. The screening unit manager insisted to pre-process the data within the screening unit’s Local Area Network (LAN) to ensure data protection and data security. The pre-processor module was executed in Java™ 2 Runtime Environment (JRE), standard edition (build 1.4.2_04-b05) version 1.4.2_04. The pre-processor module was designed based on object oriented concepts. This assisted in easily detaching the module from the JAABS algorithm and ported to the screening unit’s LAN. The testing of JAABS algorithm on the simulated dataset ensured that without any modifications to the code, the pre-processing module did not encounter difficulties when presented with the real-time dataset.

Dataset

Prior to data extraction, the Crystal Reports (CR) template was first tested with a test run. Refer to Figure 35 for the condition used to select client Sx number in the “Select Expert” dialog window, an arbitrary dataset comprising women’s record between 100001 to 250000 (Sx number) was used to extract a test set. After manually checking, it was found that the data pertaining to the period prior to 1995 was not complete (especially clinic code) and hence could not be used in the pre-processor module, so the strategy was altered and instead of using the Sx number as the selection criteria, women episode’s date of first offered appointment was selected (Figure 36).
The template included all previously screened women’s records between the dates 01 January 1995 to 01 January 2008.

The CR template for this extraction used the same data structure as depicted in Figure 37. The total number of records generated by this data extracts using the final version of the CR template was 540,539. This extract in text form was used as the input file for the data pre-processing module. The programme took 227 minutes to complete the process of generating new attributes, filtering and sorting for the whole record extract previously generated by the CR template.
This module identified episodes with missing data and removed them from the study. In total 2% (9,799) were removed as records with missing data (Table 6). It further deleted almost 3% (15,778) of the total records due to duplicate entries. The valid records constituted 86% (159,405) of the extracted dataset on an average each record had 3.2 episodes.

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total valid women’s record</td>
<td>159,405</td>
</tr>
<tr>
<td>Number of records deleted due to multiple entries</td>
<td>15,778</td>
</tr>
<tr>
<td>Records with missing values</td>
<td>9,799</td>
</tr>
<tr>
<td>CR template output records</td>
<td>540,539</td>
</tr>
</tbody>
</table>

Table 6 - Records breakdown details after data pre-processor module

The pie chart (Figure 38) shows the actual layout of the input dataset after classification by the pre-processor module. These valid records were associated with demographic details and then categorised based on the number of episodes each women’s record contained. Based on this classification, eight episode groups were formed. Episode-wise details are listed in Table 7. The eighth episode group had only 16 records and was too small for training and testing, hence has been omitted from further analyses.
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

Table 7 depicts the spread of data for each episode. The highest number of records was reached for the fourth episode. The first to fifth episodes had an average of 31,000 records. For the remaining episodes (6th, 7th and 8th) the average is only 800 records, this might have strong impact on the actual prediction capacity of the JAABS algorithm for these episodes.

<table>
<thead>
<tr>
<th>Episode number</th>
<th>Total records</th>
<th>Train set</th>
<th>Valid set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode 1</td>
<td>23270</td>
<td>4646</td>
<td>4708</td>
<td>13916</td>
</tr>
<tr>
<td>Episode 2</td>
<td>33765</td>
<td>6838</td>
<td>6734</td>
<td>20193</td>
</tr>
<tr>
<td>Episode 3</td>
<td>29497</td>
<td>5868</td>
<td>5891</td>
<td>17738</td>
</tr>
<tr>
<td>Episode 4</td>
<td>43584</td>
<td>8792</td>
<td>8839</td>
<td>25953</td>
</tr>
<tr>
<td>Episode 5</td>
<td>26669</td>
<td>5340</td>
<td>5338</td>
<td>15991</td>
</tr>
<tr>
<td>Episode 6</td>
<td>2366</td>
<td>473</td>
<td>485</td>
<td>1408</td>
</tr>
<tr>
<td>Episode 7</td>
<td>238</td>
<td>36</td>
<td>39</td>
<td>163</td>
</tr>
<tr>
<td>Episode 8</td>
<td>16</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 7 - Training, validation and test sets classified by the pre-processor module

The screening programme has recently initiated the increase in the upper age limit from 64 to 70 years. These screening batch records are in their infancy, hence the lower number of historical records for these episodes. On an average the call/recall cycle takes three years to complete a repeat of the whole population in the screening units care delivery area. This fourteen year period can be mapped to a maximum of five episodes. This was the case for the previous work carried out with the AI-ATT algorithm; it was designed for only four episodes, as a dataset was only available for those episodes. The latest increase in the age limit has certainly had a huge impact on the number of processable screening episodes. This can be verified in the current dataset, the fifth episode has 26,669 records.

This trend for episode-wise population increase is expected to continue as the number of women attending the later episodes from fifth to eighth also increases. This clearly signifies the importance of modifying the JAABS algorithm to accommodate more episodes, catering...
to the future screening unit’s requirements. The next step is to tri-furcate the dataset into training, validation, and testing sets. For each episode, a 20:20:60 proportion (for training, validation and testing respectively) was utilized to generate a complete suite of dataset (Figure 39). The training and validating sets were generated randomly by SPSS software. The remaining data (about 60%) constituted the test set. Such a classification was compulsory, ensuring a balanced approach to predicting the outcome attribute. The balanced set will contain equal amount of attendees and non-attendees. The remaining data was classified as a test set that had an unbalanced dataset i.e. the non-attendees records were only a fraction of the attendees. The future datasets which the algorithm has to predict in real-time would have an un-balanced dataset similar to the test set. The composition of the datasets thus created is depicted in Figure 39. The tri-furcated dataset was saved as separate text files with a pre-determined name for further processing by the prediction module.

![Figure 39 - Tri-furcation of dataset into training, validation and testing set for seven episodes](image-url)
Attribute selection

The next stage of the validation sequence was complete analyses of the dataset and its association with the prediction attribute i.e. screening attendance. The dataset attributes of the AI-ATT was analysed and it was found that each episode had different sets of predictor variables (Table 8). The predictor variables being categorical were analysed through parameters such as Lambda, Uncertainty, Phi (Φ), Crammer’s V and Contingency (confidence level at 95%). These tests were conducted on the probable predictor’s (independent variable) association to the screening attendance (dependent variable). Table 9 summarises the degrees of association of the independent variable to the screening attendance (highest value among the tests). The AI-ATT predictor attributes were tested on the new dataset. Upon analysing, the results indicated that some form of association, however weak can contribute to better classifying of the outcome.

<table>
<thead>
<tr>
<th>Episode numbers</th>
<th>ANNP (Post annum)</th>
<th>RBFN (Post annum)</th>
<th>ANNP (Townsend reference)</th>
<th>RBFN (Townsend reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPISODE 1</td>
<td>Age band-1</td>
<td>Age band-1</td>
<td>Age band-1</td>
<td>Age band-1</td>
</tr>
<tr>
<td></td>
<td>Type bin-1</td>
<td>Type bin-1</td>
<td>Type bin-1</td>
<td>Type bin-1</td>
</tr>
<tr>
<td>Type bin-2</td>
<td>Att bin-1</td>
<td>Cancer-1</td>
<td>Att bin-1</td>
<td>Cancer-1</td>
</tr>
<tr>
<td>Post annum</td>
<td>Age band-2</td>
<td>Att bin-1</td>
<td>Cancer-1</td>
<td>FP-1</td>
</tr>
<tr>
<td>Att bin-1</td>
<td>Cancer-1</td>
<td>Type bin-2</td>
<td>FP-1</td>
<td></td>
</tr>
<tr>
<td>Age band-2</td>
<td>Post annum</td>
<td>Int slip-1</td>
<td>Num test-1</td>
<td></td>
</tr>
<tr>
<td>Cancer -2</td>
<td>Type bin-2</td>
<td>Age band-2</td>
<td>Age band-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Num test-1</td>
<td>Post annum</td>
<td>Int slip-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FP-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPISODE 2</td>
<td>Att bin-2</td>
<td>FP-2</td>
<td>Cancer-2</td>
<td>Num test-2</td>
</tr>
<tr>
<td>Cancer-2</td>
<td>Hist can-2</td>
<td>Att bin-2</td>
<td>Att bin-2</td>
<td></td>
</tr>
<tr>
<td>Hist FP-2</td>
<td>Att bin-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age band-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPISODE 3</td>
<td>Att bin-3</td>
<td>Att bin-3</td>
<td>Type bin-4</td>
<td>Att bin-3</td>
</tr>
<tr>
<td>Age band-4</td>
<td>Cancer-3</td>
<td>Att bin-3</td>
<td>Cancer-3</td>
<td></td>
</tr>
<tr>
<td>FP-3</td>
<td>Age band-4</td>
<td>Cancer-3</td>
<td>Hist FP-3</td>
<td></td>
</tr>
<tr>
<td>Hist FP-3</td>
<td>FP-3</td>
<td>Age band-4</td>
<td>Age band-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int slip-3</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 - Predictor attributes for AI-ATT algorithm

It is to be noticed that, episodes 6 to 8 have lesser records for training, validating and testing and they were not given more weightage when deciding the predictor variables for JAABS. The rest of the episodes 2 to 5 indicated that all the variables have varying degree of association with the screening attendance. For episode 1, the same set of attributes as those of AI-ATT were used for current validation on JAABS. The new predictor variable in the form of screening distance was also tested for its association with the screening attendance attribute. For almost all the episodes varying degrees of associations were acknowledged.
Hence further detailed analysis was crucial to commit to the decision, whether to include these attributes as predictors or not.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Epi1</th>
<th>Epi2</th>
<th>Epi3</th>
<th>Epi4</th>
<th>Epi5</th>
<th>Epi6</th>
<th>Epi7</th>
<th>Epi8</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>TS</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>AttBin</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>NumTest</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Cancer</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>FP</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>HFP</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>HC</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>AttTypeBin</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>AgeBand</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Slip</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>ScrDist</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

- ☐: Variables used in AI-ATT
- ☐: Association more than 0.2
- ☐: Association more than 0.1 and less than 0.2
- ☐: Association more than 0.0 and less than 0.1
- ☐: Association less than 0.0

Table 9 - Variables association level with screening attendance for the new dataset

These issues were fundamental and are to be resolved before the entire dataset can be validated. The second episode was selected for trying out different strategies that may be adopted for the JAABS algorithm. The first part of such a strategy will decide what predictor attributes will be employed for JAABS algorithm. The other strategy will verify the validity of using screening distance as one of the predictor attributes.

**JAABS vs AI-ATT attributes**

The dataset for episode two was employed for training, validating and testing the JAABS algorithm on two different set of predictor attributes. One set consisted of variables employed in AI-ATT (Table 8). The second set contained the entire predictor variables (post annum-PA, Townsend reference-TS, attendance of previous episode-Att Bin, number of tests conducted in previous episode-Num test, cancer detection in previous episode-Cancer, false positive in previous episode-FP, history of false positive-HFP, history of cancer detected in earlier episodes-HC, invitation type for current episode-Att Type Bin, age band for the current episode-Age band, number of days in difference with actual attendance to screening appointment date-Slip and screening distance travelled by the women from her
residence to screening unit - Scr Dist). Since the degrees of association for second episode was similar to the episodes three to five and the total number of records for the second episode were approximately equal to the rest of the episodes (third to fifth episode), hence the second episode was identified to test these strategies.

<table>
<thead>
<tr>
<th>Type</th>
<th>ACC</th>
<th>NPV</th>
<th>PPV</th>
<th>SPC</th>
<th>SEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAABS with all predictor variables</td>
<td>69.3</td>
<td>47.6</td>
<td>77.8</td>
<td>45.6</td>
<td>79.1</td>
</tr>
<tr>
<td>JAABS with AI-ATT variables</td>
<td>69.6</td>
<td>48</td>
<td>78.1</td>
<td>46</td>
<td>79.4</td>
</tr>
<tr>
<td>JAABS with all predictor variables</td>
<td>69</td>
<td>30.2</td>
<td>81.3</td>
<td>33.9</td>
<td>78.6</td>
</tr>
<tr>
<td>JAABS with AI-ATT variables</td>
<td>71.9</td>
<td>32.7</td>
<td>83</td>
<td>35.2</td>
<td>81.4</td>
</tr>
<tr>
<td>JAABS with all predictor variables</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>JAABS with AI-ATT variables</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 10 - Average ROC characteristics for JAABS with AI-ATT variables

The training set had 100% ROC values for both the attribute sets (Table 10). In the validation set the ROC characteristics for AI-ATT variable alone were little better (nearly 3% on accuracy and 3% on negative predictive value were reported). Where as the test set for both the set of variables had almost the same ROC characteristics. Hence the strategy of using all the predictor variables did not affect the prediction characteristics; hence this strategy was employed for the rest of the episodes, except for episode one in JAABS algorithm. This strategy had its own advantages in the form of easy coding, less maintenance and effortless extensibility of the algorithm in catering to future extension of screening population age limits. Further, the normalisation and the classifier can be coded as common modules for all the episodes.

### Screening distance as a new predictor attribute

<table>
<thead>
<tr>
<th>Set</th>
<th>Type</th>
<th>ACC</th>
<th>NPV</th>
<th>PPV</th>
<th>SPC</th>
<th>SEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>With ScrDist Variable</td>
<td>69.32</td>
<td>47.61</td>
<td>77.81</td>
<td>45.62</td>
<td>79.15</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>67.37</td>
<td>44.83</td>
<td>76.11</td>
<td>42.11</td>
<td>78.06</td>
</tr>
<tr>
<td>Valid</td>
<td>With ScrDist Variable</td>
<td>68.99</td>
<td>30.20</td>
<td>81.29</td>
<td>33.88</td>
<td>78.59</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>64.49</td>
<td>27.19</td>
<td>76.93</td>
<td>28.21</td>
<td>76.01</td>
</tr>
<tr>
<td>Train</td>
<td>With ScrDist Variable</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 11 - Screening distance test ROC values

The screening distance variable was included as a predictor in one set of tests and a second set without. The training data had been exposed to the algorithm hence had 100% ROC values. The validation set with screening distance attribute indicated a 5% increase to ACC, 3% on NPV, 4% on PPV, 5% on SPC and 2.5% on SEN. The difference between the ROC characteristics for the test set was from 2 to 3% (Table 11). This indicated that better
prediction can be achieved with screening distance variable, but further detailed studies can assist in fine tuning the JAABS algorithm.

**JAABS ROC tests**

The earlier sections described the strategy adopted for devising the constituents of the predictor attributes. This section discusses the ROC characteristics of the tests conducted on the JAABS algorithm with the identified set of predictors. The dataset created after trifurcation into training, validation and test sets were applied on the JAABS algorithm. The training and testing was conducted for ANNP and RBFN networks as separate activities. Finally the fully automated JAABS algorithm with the voting-based classifier was tested and the results are listed in this section.

**RBFN test**

The RBFN stream was presented first with the training set and then validated with the balanced validated set and finally tested with an unbalanced test set, simulating the real-time prediction data. The only difference between this test and the real time prediction set would be a lower number of women in a given batch. For episode one, the lower number of attributes (due to the lack of screening history) and the approximation capabilities the RBFN network had almost similar characteristics with the training, validation and test sets. Nearly 47% of correct prediction (CP) and 47% of False Negative (FN) were reported (Figure 40). Episode two had improved values, CP had 67% and FN decreased to 12% and False Positive (FP) at 21%. The CP, FP and FN were found to be consistent around these same values for episodes two to five. Incidentally the number of training records was higher in these episodes. This may answer the increase in accuracy of CP for these datasets. Episode six has poor predicting capabilities, the CP was at 53%, FP at 5% and FN at 42% for the test set. These figures were consistent with the training and validating set, these could be attributed to insufficient training data. Episode seven had similar values (CP 67%, FP 7% and FN 26% for validation and CP 83%, FP 3% and FN 14%) for training, but these results were inconclusive due to the lower training and testing data.
Figure 40 - RBFN stream prediction details for all episodes
ANNP test

The ANNP stream, part of the prediction algorithm was similarly tested with the trifurcated dataset (Figure 41).
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

Figure 41 - Prediction characteristics for ANNP
The training set was always found to be 100% correct on CP, FN and FP, hence it is not represented in the figure. Similar to the RBFN test values, ANNP also had poor first episode prediction characteristics. Yet the FN was less than one third (16%) of the RBFN stream. The FP was three times (16%) of the RBFN value for the testing set. The CP was the highest at 67%. The validating set had more or less similar values. Episode two and three also had typical values, but they were improving steadily. Episode four with the maximum training data had the best values of CP, FP and FN, 79%, 10% and 11% respectively. Episode five also had similar but were less than episode four values. Like RBFN stream, ANNP also had similar prediction values. The CP was around 60%, this may be attributed to the lesser training dataset. The marked difference was seen in the first episode, in comparison, ANNP was better predicting in spite of the lower number of predictor variables (ANNP-67% vs RBFN-47%).

**JAABS test**

The JAABS algorithm was tested with the complete set of episodes after appropriate training and validating. The ROC characteristics are summarised in Table 12. The algorithm’s final prediction of the screening attendance was based on a polling strategy that relies on the prediction confidence. The accuracy of the algorithm was around 68% for the first three episodes. Episode four had the maximum accuracy at 79% closely followed by the fifth episode. The accuracy of the sixth and the seventh were the lowest between 51% and 57%. The NPV was the maximum at 51% for the fifth episode. The rest of the episodes had NPV values between 41% and 47%.

<table>
<thead>
<tr>
<th>Test</th>
<th>ACC</th>
<th>NPV</th>
<th>PPV</th>
<th>SPC</th>
<th>SEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode 1</td>
<td>67.29</td>
<td>42.07</td>
<td>76.71</td>
<td>40.22</td>
<td>78.05</td>
</tr>
<tr>
<td>Episode 2</td>
<td>69.38</td>
<td>47.65</td>
<td>77.87</td>
<td>45.66</td>
<td>79.22</td>
</tr>
<tr>
<td>Episode 3</td>
<td>69.95</td>
<td>39.45</td>
<td>76.46</td>
<td>26.29</td>
<td>85.59</td>
</tr>
<tr>
<td>Episode 4</td>
<td>79.17</td>
<td>39.25</td>
<td>87.06</td>
<td>37.37</td>
<td>87.93</td>
</tr>
<tr>
<td>Episode 5</td>
<td>76.23</td>
<td>51.61</td>
<td>83.84</td>
<td>49.64</td>
<td>84.89</td>
</tr>
<tr>
<td>Episode 6</td>
<td>57.79</td>
<td>46.51</td>
<td>64.77</td>
<td>44.92</td>
<td>66.21</td>
</tr>
<tr>
<td>Episode 7</td>
<td>51.39</td>
<td>30.02</td>
<td>76.53</td>
<td>60.05</td>
<td>48.18</td>
</tr>
<tr>
<td>Average for 4 episodes</td>
<td>71.45</td>
<td>42.11</td>
<td>79.53</td>
<td>37.39</td>
<td>82.7</td>
</tr>
<tr>
<td>Average for all episodes</td>
<td>67.31</td>
<td>42.37</td>
<td>77.61</td>
<td>43.45</td>
<td>75.72</td>
</tr>
</tbody>
</table>

Table 12 - ROC characteristics for JAABS algorithm

Episode seven had the lowest NPV (30%). These lower NPVs were expected as the proportion of non-attenders was lesser in the test set. The PPV for the fourth and fifth episodes were higher between 83%-87%. The remaining episodes had values in the range of seventies, except for sixth episode; it was 64%. The specificity metric is the most important
characteristic of the ROC values. Specificity indicates the true negatives in the population; this is related to the screening non-attenders for the given test batch. Specificity was the highest for the seventh episode at 60%, but this may not be a true indicator as this episode had only 238 records in total. The next highest value was in the fifth episode at 49%. Episode one, two and sixth had values between 40% and 45%. Episode 3 and four had lower values at 26 and 37% respectively. The sensitivity was around 80% for the first four episodes, peaking at 85% for episode 3. The higher the training set of records, higher the sensitivity values. The average for the first four episodes can be used for comparing the JAABS and AI-ATT algorithms. The ROC values for AI-ATT algorithm is shown in Table 13.

<table>
<thead>
<tr>
<th></th>
<th>Test</th>
<th>ACC</th>
<th>NPV</th>
<th>PPV</th>
<th>SPC</th>
<th>SEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode 1</td>
<td></td>
<td>67.01</td>
<td>20.45</td>
<td>87.48</td>
<td>41.81</td>
<td>71.43</td>
</tr>
<tr>
<td>Episode 2</td>
<td></td>
<td>87.76</td>
<td>56.1</td>
<td>92.85</td>
<td>58.91</td>
<td>93.14</td>
</tr>
<tr>
<td>Episode 3</td>
<td></td>
<td>86.49</td>
<td>50.54</td>
<td>92.91</td>
<td>55.99</td>
<td>91.32</td>
</tr>
<tr>
<td>Episode 4</td>
<td></td>
<td>81.65</td>
<td>41.26</td>
<td>92.51</td>
<td>64.59</td>
<td>85.42</td>
</tr>
<tr>
<td>AI-ATT Avg</td>
<td></td>
<td>80.73</td>
<td>42.09</td>
<td>91.44</td>
<td>55.33</td>
<td>85.33</td>
</tr>
</tbody>
</table>

Table 13 - ROC characteristics of AI-ATT algorithm

Figure 42 shows the comparison of the average ROC values for the first four episodes for the AI-ATT and JAABS algorithm. Even though all the ROC characteristics were better for AI-ATT, the negative predictive values were almost the same at 42% for both the algorithms. Being an exploratory research and the first iteration of a prototype, JAABS has enough potential for acceptance into the routine activities of screening unit. Undoubtedly a number of upgrades have been identified. One such area for improvement is the neural net’s ROC characteristics for the new set of data. An important facet that could be part of future work is to improve the specificity of the JAABS algorithm. Figure 42 shows the difference in the average ROC values between the algorithms episode wise. The main handicap identified for the AI-ATT algorithm was the prediction capacity of the first episode. Incidentally the JAABS algorithms had better prediction values for this episode and this not only suggest that the characteristics of neural network-based prediction is working better with lower number of predictor variables but also suggests that as the input dataset evolves and when enough training sets are available would definitely assist in making decent screening attendance predictions.
Summary

Even though the tests conducted to recognize the association between the independent and dependent variables showed varied degrees of strengths. These tests for association are only an indication of a linear relationship, whereas the real problem space is multi-spatial. Hence a new strategy of exploiting all the available predictor variables was proposed and validated in the JAABS algorithm. The accuracy and sensitivity parameters exhibit the predicting capacity of the models (Table 12 and Table 13). The values are not significantly different between the two algorithms. Based on the quantitative analysis performed on the output parameters and the earlier algorithm’s (AI-ATT) ROC values, triangulation techniques were used to compare JAABS, AI-ATT and Logistic Regression (LR) models. It was found that for episodes one to four, the LR model had only 9% NPV, whereas the other two algorithms had 42% NPV for the same set of episodes. Especially for the first episode, JAABS had a better NPV over AI-ATT (42% vs 20%). From the model’s performance perspective, all these prediction characteristics were positive.

The AI model (JAABS) proposed were consistent to the earlier model’s (AI-ATT) performance and consistently out-performed traditional statistical techniques (LR). This could be attributed to the complex nature of the input variables. The knowledge creation by applying AI is not only consistent, repeatable and economical; but also ensures minimal human intervention. This is perfectly ideal for automating the whole process. These were the principal requirements for the prototype model that has been fulfilled through appropriate validation in this chapter.
Knowledge Sharing: GP’s Role

In the last chapter, the GP’s role and the proposed knowledge sharing were described in detail. This chapter identifies the importance of GP interventions and their impact on breast screening attendance. Earlier chapters described how quantitative research would be applied through a questionnaire-based study. In concurrence to this, the following discussion first covers the factors which influence the women’s breast screening attendance and details that are related to approaches that can be used to address such factors. It provides further analyses through a questionnaire-based survey conducted with Coventry, Solihull and Warwickshire GPs. The survey covered a wide range of areas related to interventions; including the favoured types of interventions, mode of message delivery, updating methods and requirement of additional resources. Combined with the information collected during the pilot study (from the GPs), this analysis would shed more light on the finer points of intervention and assist to triangulate these results.

GP intervention

Breast screening non-attendance can be attributed to semi-permanent and temporary factors. Semi-permanent factors such as ethnicity, age group, marital status, income, education and chronic conditions can be major influencing factors on the screening population (Katz et al. 2000). These factors are difficult to address and usually consume more time and resources to make a significant impact; whereas temporary factors such as employment, personal apprehensions, beliefs, knowledge and access to screening facilities can be targeted effectively to increase the screening attendance (Sin and Leger 1999, Bekker et al. 1999). Publicity campaigns, media-based advertisements, knowledge-based education drives, posters, pamphlets etc. are the primary counter measures planned and executed by the NHS to address the temporary factors affecting screening attendance. Research by Kelley has found that by bringing the screening services to the women rather than making the women come to the screening office can greatly influence the attendance rate (Day et al. 1989). The NHS has many mobile screening units under its service and they are strategically placed at public places such as shopping centres and other public frequented places for maximum effectiveness.

Types of intervention

Evidence points out that intervention through printed and audio-visual educational materials and education sessions has limited effectiveness, but when these strategies are coupled with GP interventions, they can help the non-attending women to make informed decisions on screening (Jepson et al. 2000). Primary care can play a vital role in addressing some of the important temporary factors affecting screening attendance such as personal
apprehensions, beliefs, and knowledge on breast cancer screening (Fox et al. 1991, Bekker et al. 1999). Several studies have established that primary care physician’s (GP’s) advice and recommendations are important factors that guide women to make educated decisions on screening (Sin and Leger 1999, Kelley 1999, Bennett et al. 2000, Vainio and Bianchini 2002, Day et al. 1995, Turner et al. 1994, Fox et al. 1991). Further research has identified that family physicians can considerably help to increase attendance through appropriate interventions. These can take the form of GP signed letters, one-to-one meetings, telephonic or opportunistic interventions (Turner et al. 1994). The current research work proposed to create knowledge on screening non-attendance through prediction and share it with primary care deliverers through effective knowledge sharing techniques (Baskaran et al. 2004). Such knowledge when appropriately shared can be leveraged to initiate reminders at the right time in a clinical context. Reminders, if properly utilised, could be part of a potential strategy to improve the underused preventive care services (screening) (Shea et al. 1996). The lack of research in this area prompted to further investigate the intricacies and strategies to be employed while adopting the proposed KM-based approach.

Questionnaire-based survey

The lack of evidence in this area and the recent advances in Information Technology (IT) prompted the collection of additional data to determine the best possible use of the gained knowledge, in particular data about the preferred mode of knowledge delivery to GPs, their opinion on the interventions and their perceived need for additional resources/assistance (refer to blank questionnaire in Error! Reference source not found. for more details). This research conducted a survey of the local GP surgeries which are geographically related to a local breast screening service in West Midlands, UK. All the GP surgeries (N=215) covered by the Warwickshire, Solihull and Coventry Breast Screening Service were eligible for the sample survey. Lists of GP surgeries were obtained from the area’s Primary Care trusts. The GP surgery population spread across the service can be classified by the following areas: Warwickshire (north) 38, Warwickshire (south) 48, Solihull 36, Rugby 15 and Coventry 78. In total, the questionnaires were dispatched to 215 GP surgeries. Only 30 (14%) responses were received in the first round and this was followed by a reminder letter which yielded an additional 29 (13.5%) responses (refer to Error! Reference source not found.). A total response rate of about 59 (28%) was achieved; this was slightly below average for a postal questionnaire based survey. Three questionnaires were returned back due to change of surgery address.
### Sample

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
<th>Queries answered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screening women</strong></td>
<td>Target population of the screening programme</td>
<td>Population (approx.) of women registered with GP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>List of clinical condition and its frequency</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td>Providing prior information on non-attendance through prediction</td>
<td>Usefulness of prediction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate duration for intervention</td>
</tr>
<tr>
<td><strong>Electronic Data Interchange</strong></td>
<td>Electronic exchange of information to GPs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Usefulness of paper based screening report</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of screening report and its update to women’s record</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preferred mode of future knowledge and information delivery</td>
</tr>
<tr>
<td><strong>Impact</strong></td>
<td>Impact of GPs on screening attendance</td>
<td>Probable impact on breast screening attendance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preferred intervention type and its duration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preferred staff for intervention</td>
</tr>
<tr>
<td><strong>Additional information</strong></td>
<td>Aspects related to recognition of GP intervention</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Recognition for intervention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requirement of additional resources</td>
</tr>
<tr>
<td><strong>Demography</strong></td>
<td>Information on surgery and GP</td>
<td>Sex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of partners</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surgery population</td>
</tr>
</tbody>
</table>

*Table 14-Issues covered by the questionnaire (Baskaran et al. 2006)*

### Survey

The survey was conducted in the summer of 2006 through a mailed four page questionnaire which enquired about GPs belief on the screening prediction, mode of knowledge delivery, intentions with the proposed knowledge sharing and any additional resources required to intervene (Table 14). The survey was accompanied by an executive summary of the research envisaged. The survey design and analysis was approved by Coventry University’s Research Ethics Committee and the NHS Local Research Ethics Committee (LREC). The questionnaire was piloted for its context, logic and readability with two groups of experts, one containing five academics within the university and five GPs located in the Coventry area. The GPs suggested that the covering letter, if addressed to the practice manager, would result in a better response rate (earlier it was addressed to the senior physician of the surgery). Some of the questions were found too complex and the GPs suggested splitting the questions into smaller and simpler ones for a better response. During the pilot, some of the technical (non-clinical) terms were suggested to be avoided for better understanding.
Earlier research by Bankhead et al. (2001) had suggested that intervention through GPs was useful but was found not to be a cost effective strategy. Moreover, this study was conducted before 2000 and the inferences (cost-based) may be outdated. Other observational studies and manual prompts in women’s records were found to assist GPs in encouraging non-attending women to attend a scheduled screening appointment (Sin and Leger 1999). The current survey was expected to validate four important factors pertaining to the intervention. The first assumption is that the screened women aged 50 to 70 would be visiting the GPs on a regular basis. The next hypothesis relates to the opinion of GPs on the vital part they can play in the quest for improving screening attendance through intervention. The third assumption is that GPs would prefer opportunistic intervention when compared to other types of intervention. The final hypothesis relates to the electronic version of knowledge transfer and its preference over the traditional paper-based screening results (Baskaran et al. 2006a).

### Questionnaire analysis

The preliminary analyses of the responses pointed to a bimodal distribution. Hence the responses are dichotomised as follows (refer to Table 15). The data was analysed using SPSS® (12.0) for Windows software (Hinton et al. 2004). The attributes employed in the analyses include demographic variables such as gender, practice experience, number of partners and surgery population. The other attributes relate to screening women, prediction, Electronic Data Interchange (EDI), GPs impact on attendance and information on additional resources required (refer to Error! Reference source not found. for the results frequency distribution). Multi-variate and bi-variate analyses were conducted on the

---

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Dichotomised variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening women</td>
<td>Frequency of women approaching GP surgery</td>
<td>• Often</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not at all</td>
</tr>
<tr>
<td>Prediction</td>
<td>Usefulness of prior information on non-attendance</td>
<td>• useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not useful</td>
</tr>
<tr>
<td>EDI</td>
<td>Usefulness of paper based screening results</td>
<td>• useful</td>
</tr>
<tr>
<td></td>
<td>Helpfulness of electronic mode of results delivery</td>
<td>• preferred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not preferred</td>
</tr>
<tr>
<td>Impact</td>
<td>Impact of GPs intervention on screening attendance</td>
<td>• weak impact</td>
</tr>
<tr>
<td></td>
<td>Preferred staff to conduct intervention</td>
<td>• preferred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not preferred</td>
</tr>
<tr>
<td>Additional info.</td>
<td>Recognition recommended</td>
<td>• strongly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• weakly</td>
</tr>
<tr>
<td></td>
<td>Requirement of additional resources</td>
<td>• less</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• more</td>
</tr>
</tbody>
</table>

Table 15- Dichotomisation of the variables categorised by questionnaire subsections
attributes to estimate and experiment the influence of the covariates on the various sets of
dichotomised variables. An exhaustive literature review within the study domain was
conducted to classify the appropriate attributes for this study. In addition, the independent
variables based on their importance were incorporated in the logistic regression model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Usefulness of screening results</th>
<th>Usefulness of prediction</th>
<th>Helpfulness of electronic results</th>
<th>GP intervention has strong impact</th>
<th>GPs ability to intervene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>p Value</td>
<td>Frequency</td>
<td>p Value</td>
<td>Frequency</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>100%</td>
<td>0.131</td>
<td>88%</td>
<td>0.41</td>
<td>68%</td>
</tr>
<tr>
<td>Female</td>
<td>89%</td>
<td>0.41</td>
<td>79%</td>
<td>0.008</td>
<td>95%</td>
</tr>
<tr>
<td>Practice experience</td>
<td>0.182</td>
<td>0.436</td>
<td>0.053</td>
<td>0.583</td>
<td>0.663</td>
</tr>
<tr>
<td>&lt;10 years</td>
<td>100%</td>
<td>0.182</td>
<td>88%</td>
<td>0.436</td>
<td>100%</td>
</tr>
<tr>
<td>&gt;11 years</td>
<td>90%</td>
<td>0.436</td>
<td>78%</td>
<td>0.053</td>
<td>81%</td>
</tr>
<tr>
<td>Partners</td>
<td>0.449</td>
<td>0.215</td>
<td>0.917</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>&gt;2</td>
<td>91%</td>
<td>0.449</td>
<td>71%</td>
<td>0.215</td>
<td>91%</td>
</tr>
<tr>
<td>&lt;2</td>
<td>96%</td>
<td>0.215</td>
<td>80%</td>
<td>0.917</td>
<td>77%</td>
</tr>
<tr>
<td>Registered population</td>
<td>0.943</td>
<td>0.171</td>
<td>0.693</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>&gt;5000</td>
<td>93%</td>
<td>0.943</td>
<td>82%</td>
<td>0.171</td>
<td>93%</td>
</tr>
<tr>
<td>&lt;5000</td>
<td>93%</td>
<td>0.943</td>
<td>81%</td>
<td>0.171</td>
<td>81%</td>
</tr>
</tbody>
</table>

Table 16 - Cross tab analyses on screening, prediction, EDI and intervention against demographic details

The multivariate and bi-variate analyses considered significant values of <0.05. Odds Ratio (OR) and Confidence Interval (CI) were also performed. 38 (67%) of the respondents were female GPs. 41 (73%) of the surgeries had four or less partners. Nearly 25 (49%) of the GPs who responded had more than twenty one or more years of practice and 18 (35%) had experience between eleven and twenty years. Nine respondents did not provide their practice experience.

Screening women

The majority of responding surgeries 46 (81%) had a registered population of less than 10,000 and the remaining surgeries had populations of more than 10,000. The national average for registered populations in UK surgeries is 8,400 (The Information Centre 2007), whereas the average registered population for each practice within Coventry Primary Care Trust PCT was 5000, Warwickshire PCT 4900 and for Solihull PCT was 6500 (The
Information Centre 2007). Most of the respondent surgeries 40 (76%) had 400 or more women (aged between 50 and 70) registered with them.

The majority of the respondents reported that women aged between 50 and 70 approached the surgery more often for Muscular 46 (92%), Diabetes 44 (88%), Arthritis 44 (88%) and Hypertension related clinical conditions 47 (94%). There was 10% missing data among the respondents on this question. Nearly half, 26 (49%) of the responding GPs visited women (aged between 50 and 70) at home for various health related services, 24 (37%) of nurses made less than 5 visits per month to the same aged women. Special nurses made 15 (30%) visits to less than 5 women per month for the same age group. The association between the surgery demography and screening results usefulness has no statistical significance due to higher \( p \) values, Table 16).

**Prediction**

41 (82%) of the GPs agreed that it would be very useful to provide prior information through prediction about non-attending women, so that they can intervene and reduce breast screening non-attendance. The majority of the GPs, 50 (93%) are of the opinion that about thirty days (in advance) would be an ideal period to initiate intervention on the predicted non-attending women. The usefulness of the prediction was not significantly associated with the surgery demographics, but those surgeries with less than two partners were associated (to a certain extent) with the GPs who found prediction useful (Table 16).

**Electronic Data Interchange (EDI)**

Most of the GPs 55 (93%) believed that it is very useful to receive screening results (in paper format instead of electronic) to provide better health care to women. 52 (92%) GPs agree that currently they are receiving screening results in paper format from the screening service. Bivariate analyses of EDI and the gender of GPs who responded to the questionnaire revealed that female GPs are more likely to accept electronic results (95% vs 68%, \( p=0.008 \)). Almost the same percentage agreed that these results, after being received, are always updated into the women’s records. Out of these GPs, 33 (83%) mentioned that the results are manually read from the report and added into the women’s records. The remaining surgeries answered that they are scanning the screening result document and storing it as a computerised file along with the woman’s record in their local database. Even though bivariate analyses did not show any significance that EDI based results are more favoured by all GPs, further analyses indicated that GPs with less than ten years of experience are more likely to accept electronic based screening results (100% vs 81%, \( p=0.053 \)). 51 (86%) of the GPs who responded to the questionnaire were of the opinion that it would be very helpful to receive the electronic version of the screening results.

They also deliberate further that they (85%) preferred an automatic updating of results which would be sent directly to their database. The least (70%) preferred method would be a manual updating method utilising a Compact Disc (CD) with results burnt on it and sent to the surgeries through the postal services. Incidentally, the GPs’ preferences were split equally on supervised and unsupervised updating; in this case the results would be sent to
GP intervention

Three quarters 45 (76%) of the GPs suggested that interventions initiated by them have a strong impact on increasing screening attendance. Nearly half 20 (49%) of the GPs who responded chose (one to one type) opportunistic intervention as the most suitable form of intervention. Just under a third (17) believed that making an intervention by post to the non-attending women would be more suitable. 13 (23%) preferred a telephone-based intervention. Of those who preferred opportunistic intervention, almost all 18 (98%) consider that the intervention would not take more than ten minutes. While selecting the preferred surgery staff for intervention, 45 (88%) respondents believed that nurses were appropriate to conduct an intervention in the surgery and 35 (75%) considered the doctor was more suited for intervention. In addition, 26 (60%) of the GPs indicated that appointment booking assistants were preferred interveners. Surgery demography was not significantly associated with the GPs who thought that there is strong impact through intervention (Table 16).

Additional information on interventions

Three quarters of the GPs 43 (73%) responded that they have the ability to intervene if provided with the knowledge of non-attendance. Bivariate analyses show significant association between the GPs who said that their ability to intervene can make a positive impact in increasing uptake and the number of women population registered with them (86% vs 61%, p=0.035) (Table 16). The majority of GPs 40 (80%) wanted recognition for the efforts put in by the surgery for increasing the screening uptake. Out of the GPs who believed that their efforts are to be recognised also regarded that appropriate awards 16 (26%) and proper appreciation 29 (48%) were recommended by them. Almost 39 65% of the GPs thought that their efforts should be compensated through remuneration. Although the difference did not reach statistical significance, surgeries with more than two partners were more likely to accept that GPs interventions have impact to increase screening attendance than those surgeries with only one physician (82% vs. 62%, p=0.082) (Table 16). The GPs responding to this section also indicated that additional assistance is needed in the form of patient directed leaflets on breast cancer in English 43 (72%), leaflets in other languages 39 (65%) and more training to GP staff on breast screening (58%) in order to make better interventions for increasing breast screening attendance. Fifty percent 35 (50%) of GPs also expected more assistance in the form of videos on breast screening and GP info-packs.

Questionnaire results

As part of a wider research, this survey was directed at finding the associations of various attributes with screening results, predictions and GP interventions. Many of the respondents were female GPs; this is of some importance as this questionnaire was not directed specifically to any gender. The practice manager to whom the questionnaire letter...
was addressed might have expected female GPs would be more interested to answer the questionnaire yet there were sizeable practices 22 (39%) with two or less physicians practising. The national average of two or less partners in a surgery is around 20% (RCGP 2005). The number of female GPs in UK has increased from 23% to 35% (RCGP 2005); all these factors might be the reasons for more female GP responses. The number of women aged 50 to 70 registered with a GP was expected to be evenly distributed according to the GPs who participated in the questionnaire piloting phase. Although the respondents had indicated that the majority had more than 400 women registered with them, the data collated for the same age group from the Royal College of General Practitioners (RCGP) information sheet has indicated that, on average, the number of women registered with UK GPs is approximately 200 (The Information Centre 2007, RCGP 2005). The figures collected in this questionnaire, which is double the national average, is a clear indication that this sample is not truly representative, yet there is a sizeable women population in the study area who are eligible for breast screening and who are also registered with GPs. Increasing consultation rates from 5 to 7 consultations per person year) from 1995 to 2006, and a rise with age (Table 15) indicates that GPs do have the opportunity to address screening attendance through opportunistic intervention (Figure 58). Jepson et al. (2000) considered that educational home visits to be one of the effective interventions. The NHS information centre’s study indicates that the number of GP home visits halved (from 9% to 4%) (The Information Centre 2007) reducing the opportunity to intervene in a non-surgery environment. On the other hand telephone-based consultations have trebled (from 3% to 10%). At the same time provided additional possibilities for making opportunistic intervention to increase screening attendance.

**Figure 43 - Consultation rate per person year trend (1995-2006) for age group between 50 and 69**

The study also reiterated that the GPs agree that the majority of the women who are aged between 50 and 70 approach the GPs for almost all the listed conditions such as muscular/skeletal disease, diabetes, arthritis and hypertension. Nearly 15% of the respondents also indicated that this statistic was too difficult to collate or requires more efforts to identify the numbers. Some of the GPs also reported that the women of the same
age also visit more often for minor illnesses such as infections, stress related problems and upper respiratory tract infections, to specify a few. Earlier studies by Sin and Leger (1999) indicated that clinical staff often makes house calls / visits to those women in the age group 50-70. It is imperative that the current research (reported in this paper) identifies the staff making house calls in the primary healthcare team. More information pertaining to the frequency of their house visits would be helpful to recognize the possibilities for making interventions in a non-surgery type environment. Nearly half of the doctors make at least 6 or more visits (per month), 15% of nurses make 6 or more house calls and 11% of special nurses make 6 or more home visits. Out of these, the doctors were the preferred choice of clinical staff to make such kind of house call. The house calls are expected to be conducted in a more relaxed environment and this could yield a good doctor-patient relationship (Hinton et al. 2004).

This creates a unique opportunity for the doctor to have a personal rapport with their patients and hence is more suitable for making an opportunistic intervention to address the screening non-attendance issue (Sin and Leger 1999). Further analyses identified that the GPs who responded positively about the impact of GP intervention can be predicted by the clinical staff house visit calls attribute ($R=0.494$ and $\chi^2=9.519$, $p<0.02$). How ever insignificant, one GP mentioned an interest in intervening but is handicapped through lack of time, this shows that some GPs may have intent to intervene but are restricted due to lack of resources. Predicting women’s breast screening attendance and sharing the knowledge is new concept. The response of the GPs regarding the usefulness of prior knowledge on non-attendance (so that they can intervene) is an important issue. Most of the GPs considered that prediction would be useful. This had strengthened the assumption that GPs themselves believe that they accept the idea of receiving the $a$-priori knowledge of non-attendance. The attributes for predicting the GPs acceptance of prior knowledge is dependent on those who also believe that GPs have the ability to make a positive impact through intervention to increase breast screening uptake ($R=0.450$ and $\chi^2=11.816$, $p<0.001$).

An important aspect of this research is to identify at what point the AI-based prediction should be run and sent to GPs. This should neither be too short - which would result in limited time available for an opportunistic intervention to take place - nor too long to the point of diluting the fervour of intervention. The respondents overwhelmingly agreed that about four weeks in advance would be appropriate for initiating intervention. These four weeks are well within the screening process schedule. A few (less than 10%) responded positively for thirty or more days prior to the date of screening. A two-pronged strategy recommended by this project (Baskaran et al. 2005) aims not only to gain advantage of prediction knowledge but also the knowledge gained through the post-screening results on non-attendance through appropriate intervention. The said intervention may happen over time as it would be conducted after the completion of the screening cycle. All these factors warranted further evidence on how the screening results are received by the GPs, their updating frequency, mode, etc. At present, the screening centre mails the screening result copy to the women and to their GPs. The screening results are in paper format and are filed with the woman’s clinical record at the GP surgery. The respondents have unanimously agreed that the screening result (in paper format) is very useful in providing better healthcare to women. During the questionnaire piloting phase; there were some apprehensions about the usefulness of the screening results and possibility of the results not being handled appropriately.
Hence the GPs were also asked whether they obtained the screening results and most of them concurred that they receive the results, and the same number of respondents recognised that those results were always updated to the women’s records. Out of the sample who responded, 3 (5%) answered “No” and 2 (3%) answered “Did not know” as to whether the results were received, so there is still scope for improving the process of screening results delivery. In the UK, primary care has been computerised and the majority of surgeries are utilising computers to assist in their care delivery (Protti 2005). This was also reflected in the responses to the question on the mode of surgery updating. Twelve percent (12%) of the GPs suggested that the paper format results are scanned as files and stored directly in their database, and half of the GPs responded that the results were manually updated as text with ‘Read Code’ into the women’s records by either the GP themselves or by the practice manager or nurse or, in some instances, by the data clerk. Computers play a vital role in the surgeries and provide assistance for laboratories to send the results directly to the GPs in electronic format (Protti 2005); most of the GPs (86%) agreed that it will also be helpful to get the screening results in electronic format.

Other screening services utilise Compact Disc (CD) for transferring the results of screening to the GPs. This format facilitated database updates; this not only saved time but also avoided errors while updating, still the transfer to the computer system needed manual involvement. 27 (45%) of the GPs did not answer this option. Supervised updating of results can be accomplished when the results are being delivered such as an email message. Supervised updating mode response was identified as a predictor for GPs who agree that prediction is useful for increasing attendance ($R=0.362$ and $\chi^2=4.221, p<0.04$). This question was also not answered by as many as half of the GPs who responded. Almost 85% agreed that the results, if sent directly to their database, would be the most preferred option. GP intervention is the core issue of this questionnaire. The GPs should have faith that the interventions made by them have strong impact on increasing screening attendance (Taylor et al. 2003). This was reflected by the positive responses 45 (75%). Earlier research by Richards et al (2001) suggested that opportunistic intervention was more economical and effective (Bell et al. 1999). Forty seven percent 27 (47%) of the GPs concurred with the opportunistic (one-to-one) type of intervention as preferred. To our surprise, the traditionally preferred letter-based intervention (Turner et al. 1994, Majeed et al. 1997) for screening was also preferred by 17 (30%) of the responding GP. One of the earlier studies identified that even the appointment booking assistant, if trained appropriately, can play a positive role in increasing screening attendance (Atri et al. 1997) and this idea has been strengthened by the GPs who responded to our questionnaire: they preferred appointment booking assistants (60%) as suitable intervening staff. Healthcare assistants were also suggested by one of the GPs as a preferred person to conduct the intervention.

Multi-variate analyses of preferred intervention personnel (nurses and booking assistants) attributes indicated with statistical significance ($R=0.507$ and $\chi^2=10.388, p<0.006$) that they were positive predictors in identifying the GPs who felt that prediction was useful. The predictive accuracy increased when the doctors were also included in the set of attributes ($R=0.578$ and $\chi^2=13.429, p<0.004$); this increased the prediction’s statistical significance. One of the GPs strongly responded that there were no resources available to allow involvement in any kind of intervention. The knowledge of non-attendance has been perceived to make a positive impact through intervention to increase breast screening
uptake (Sin and Leger 1999). Nearly 26% of the GPs either answered “No” or “Did not know” for this query. The GPs who thought that their efforts for making intervention has to be recognised have also preferred remuneration 39 (65%) as the most recommended form of recognition followed by appreciation 29 (48%) and awards 16 (26%).

The qualitative study conducted during the piloting phase added proof to the findings of the earlier study by Sin and Leger (1999); both had indicated that remuneration-based recognition was the most preferred method. However, triangulation of the results to the questionnaire respondents revealed a different view; appreciation and awards were preferred by some GPs. These questions were aimed to find two decisive factors. 1) Whether the GPs felt the need to be recognised for their efforts to make intervention? 2) Do they feel that intervention was part of the general care delivery? The questionnaire respondents clearly indicate that some form of recognition is vital for making the intervention-based increase to screening uptake and GPs were of the opinion that they cannot do the intervention as part of general care delivery due to the lack of resources. Analysing the questionnaire, some GPs asked for additional resources (within their surgery) to make an impact on screening attendance. The questionnaire response also supported this aspect as about three quarters of the GPs (74%) recommended all the listed options in the questionnaire. In order of preference, the GPs thought that they require more patient-directed leaflets on breast cancer in English and in other languages, training of GP staff, GP info-packs and videos about breast screening. Even though most of the above mentioned that assistance is being provided by the NHS, still there is scope either to increase the level of assistance or to cover more surgeries with said assistance.

There were certain limitations to the current study. This questionnaire was aimed at small geographically related GPs and had less than average responses; hence it might not accurately reflect the entire GP population’s views in the UK. This was the reason that most confidence intervals had wide ranges. Interventions were assumed to improve static screening uptake figures. Moreover women are encouraged to make informed decisions about attending a breast screening episode (Turner et al. 1994). Hence interventions alone may not be the contributing factor for improving the static uptake figures. This research relies heavily on the success of the proposed NHS-CfH’s IT project and the continued use of computers in the primary care domain. Prediction through AI is not 100% accurate and errors are expected in the form of false positives and false negatives. Some of the questions had very low response rates which might warrant a more detailed study on the specifics.

The questionnaire has strongly suggested that there is a high probability that screening aged women visit the surgery for various health related reasons and hence provide an opportunity for the GPs to intervene. This study clearly suggests that GPs strongly believe that they can play a vital role in increasing the screening uptake. Owing to the fact that, in the UK, the primary care deliverers have been transformed into small business unit type of organisations, this had a marked impact on their way of healthcare delivery. GPs believe that all the time they have is to be spent in monetary oriented clinical services. This may have increased the overall efficiency of the surgeries and their care delivery effectiveness, but resulted in the non-remunerative and indirect primary care related services, such as screening, taking a back seat. Some screening services such as cervical screening are offered by the surgeries and they are compensated for the services rendered to the women by the
NHS. Similar compensations are recommended to make the proposed breast screening attendance based interventions a success.

The UK NHS has embarked on its most ambitious IT project and it is building the infrastructure for providing automated Electronic Data Interchange (EDI) across its entire clinical domain. Our study hypothesised that the utilisation of the infrastructure developed by the NHS Connecting for Health (CfH) can be effectively used to deliver the predicted screening attendance knowledge through Health Level 7 (HL7)-based messages (Baskaran et al. 2004, 2006a, 2006b, Protti 2005). Continuous improvements to computer networks and IT infrastructure are the norm of modern healthcare. This would not only be highly beneficial for increasing the uptake but would also be economical. The GPs have provided this questionnaire with an overwhelming positive response on their willingness to automate the mode of message delivery.

Survey summary

All these factors and the response to the questionnaire have suggested the following. If such kind of interventions (based on prediction or post-screening results) are to be made effective, some form of compensation to cover economical overheads in surgeries is essential, otherwise these initiatives are destined to fail. The same information was reflected during the qualitative study conducted earlier. The responses did not provide a clear picture on the preference of opportunistic intervention over the traditional telephone- or letter-based interventions. More focused studies in this area can shed light on these aspects. Earlier studies on computer generated reminders for GP interventions had suggested an increase in screening uptake (O’Connor et al. 1998, Shea et al. 1996). Hence an electronic version of knowledge transfer to flag reminders is highly favoured by the GPs. The questionnaire results (on the mode of knowledge sharing) concurred with the findings of the earlier qualitative study on the same issue. This augmented well with the research strategy adopted for this study; the triangulation of the results provided a benchmark which supported the researcher’s beliefs on GP interventions (to increase screening attendance). All the above inferences are very encouraging and provide the required evidence that GP interventions are effective and can improve screening attendance further, it reveals that computer initiated interventions would be economical and effective.
Conclusion and Future Work

This project set out to answer the research question and deliver the objectives mentioned in chapter 1. This project not only confirms that breast screening attendance can be predicted through an automated software solution but also can be leveraged to increase screening attendance by employing emerging KM tools and techniques. This work has also proved that such techniques can be implemented using current technologies available at the NHS' disposal.

A prototype model (encompassing the JAABS algorithm) was designed and implemented on Open Source technology as part of the project. This prototype model has become the first iteration in preparing a complete software-based solution for addressing the breast screening attendance. The project also validated the prototype through training and testing with thirteen years of historical screening data from the Warwickshire, Solihull and Coventry Breast Screening (WSCBS) unit. The first hypothesis of this project “that it is possible to design and implement an Open Source-based prediction algorithm prototype for breast screening attendance” has been tested and validated using the current datasets. The results were then triangulated with the AI-ATT algorithm’s test results.

In addition, the proposed knowledge transfer was accomplished using appropriate messaging standards (Health Level 7). Further, a detailed description on the design and implementation of the new message suite (intended exclusively for sharing knowledge with the primary care deliverers) was provided. The new Breast Screening Attendance Messaging Protocol (BSAMP) that can be implemented on the National Care Records Service (NCRS) being pursued by the NHS was depicted earlier. The second hypothesis “it is possible to design a messaging platform for knowledge creation (through prediction) and sharing model (knowledge/information exchange architecture) with available technology and infrastructure” has also been tested and a workable model has been highlighted earlier. Implementing the BSAMP on NCRS would require no additional investments. This lends support to the research proposal’s ability to be integrated with minimal effort from screening staff and GPs. Such initiatives would not only be economical but could be rapidly deployed. A survey of the physicians for the WSCBS area confirmed that GP interventions are feasible through electronic prompts and such interventions will undoubtedly improve breast screening attendance.

The research also highlighted the advantages of applying both qualitative and quantitative research methodology in combination. The synchronisation of these methodologies not only supported each other but also provided a baseline for comparison using triangulation techniques. Furthermore, the research approach adopted for this highlighted the importance of conducting qualitative research. The qualitative research strategy also clarified the research direction taken and offered support for further quantitative research at a later stage. This work also presented the need for a balanced research methodology for maximising...
deliverables. The questionnaire tested the last hypothesis “it is feasible for GPs to make interventions to increase breast screening attendance through electronic prompt” and the results were discussed.

Conceptual development

One of the original aims of the current research was to employ KM and its associated technologies to increase the breast cancer screening attendance. A true KM-based project has to address all the aspects of knowledge and its components. This section explains the development of KM-related concepts so that an original contribution can be made to the body of new knowledge. The following section has been structured based on the three core components detailed earlier.

Concept development - Knowledge management

The first phase focussed on a critical literature review which highlighted the clear need and scope for redefining the very essence of KM. This prompted a new examination of how knowledge needs to be viewed. In particular, the differentiation of explicit knowledge and information was of paramount importance when applying KM to the healthcare field. The concept and efficacy of KM within healthcare is not always understood and the notion of such critical terms as “information” and “explicit knowledge” are often (incorrectly) regarded as being interchangeable. We elaborate later on this concept through an easy to understand clinical example. The view proposed as first and second generation information is a new concept and clearly differentiates the differences between these terms. Equipped with this new understanding, even novices can clearly comprehend these fundamental elements and avoid using them as misnomers. This ensures that knowledge and its components will be properly addressed in both clinical and non-clinical domains.

Concept development - Artificial Intelligence (AI)

This new concept refers to the Artificial Intelligence algorithm which was developed for the first time on an open source platform for the NHS breast screening programme. This platform was pursued due to fiscal limitations and commercial licensing difficulties. Even though applying an AI-based algorithm for breast screening prediction was not new (Arochena, 2003), the proposed JAABS algorithm is different from its predecessor (AI-ATT). The primary difference is that the previous algorithm worked on a commercial software platform and used a number of disjointed software tools for prediction. In addition, it depended exclusively on the developer’s expertise to connect these tools and work as a comprehensive prediction tool. The new JAABS algorithm is not only built on open source technology but was completely automated, requiring no human involvement for completing the breast screening attendance prediction.

In addition, this concept also identified a new attribute for improving the prediction efficiency of the algorithm. This attribute was developed based on a novel approach which employs postcodes (of the screening unit and screening women). These postcodes were
transposed to their corresponding latitude and longitudinal values. Using these values, the
distance between the two locations has been calculated and consolidated as a categorical
variable and input to the AI network for breast screening attendance prediction.

Concept development - Health Informatics

The last conceptual development relates to the health informatics aspect of this work. The
NHS has initiated the design of HL7-based messages for a complete gamut of messaging
domains. This message development process is undergoing a phased delivery and is yet to
cover the breast screening domain. The current research not only created a set of messages
using HL7 version 3 protocols but also proposed a new message design template using the
very tools HL7 UK recommends for developing them. This unique template corresponds to
knowledge sharing both at the pre- and post-screening stages. The concept of employing
HL7 messages for knowledge sharing to address breast screening attendance is the first of its
kind. This exercise has proved that developing HL7 messages, for even a limited scope, is
possible with currently available software tools.

Original contribution

In this project, several original contributions have been made to the body of knowledge
associated with healthcare. The novel concept of screening attendance prediction in an Open
Source environment was new to the breast screening domain. In addition, an innovative
prototype was constructed to investigate the validity of the idea proposed through
independent research and analyses. Eardley and Czerwinski (2007) in their paper had
highlighted the importance of KM in healthcare and how the NHS has embraced and
adopted these tools, systems, and strategies for many KM initiatives. This research draws its
strength from such KM tools and techniques. This work is also one such initiative
addressing the NHS’ breast screening attendance through efficient KM methodologies.

This research has yielded many ideas; in particular, a fundamentally new concept of
Knowledge Management (KM) and its constituents were identified. Such radical new
concepts are going to revolutionise how future knowledge specialists in health domain are
going to view and apply KM for better care delivery. A new, simple and ready to use KM
framework and model were proposed and this project used such a model as part of its
validation. Such new concepts and models are not limited to healthcare domain only. These
models can be easily adapted to any business domain with minimum amendments. This
research was funded by the NHS Cancer Screening Programmes. For the first time, this
project had compared the most common healthcare messaging standards employed in UK.
A comparison table, listing the differences between the Electronic Data Interchange For
Administration, Commerce and Transport (EDIFACT) and Health Level 7 (HL7) standards
is presented in an earlier section of this project.

As this project work was part of a larger research group’s activities, it extended the
previously completed doctoral research through continuous and focussed research. This
type of continued research not only adds value to the cancer care and breast screening; it
also identifies new ways to further the knowledge boundary through iterative, self-
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

sustaining, and focussed research cycles. In doing so, it also validated the previous research with updated set of data thus gaining more credibility to both the initiatives (AI-ATT and JAABS). The empirical work completed on KM, AI and healthcare messaging front has not been experimented in such a combination before for increasing screening attendance. This work ideally would be the first of its kind in the breast screening domain. This work would attract further research for continuously improving the care delivery process related to breast cancer at the national level.

From the context of multi-faceted research, this project is highly interdisciplinary in nature. This work that had captured emerging technologies such as HI, AI and KM, creates a synergy in these fields for addressing breast screening issues. All of these said new contributions were found worthy of publishing in international conference proceedings and journal publications. Such peer reviewed publications has certainly augmented the value and added the quality to this project. A summary of some these original contributions are listed in Table 17.

<table>
<thead>
<tr>
<th>Competency</th>
<th>Corresponding project work</th>
</tr>
</thead>
</table>
| Evidence of an original investigation | • prediction knowledge sharing  
• new predictor variable |
| Competence in independent research | • KM in healthcare  
• neural network prototype implementation  
• questionnaire-based survey on GP interventions  
• HL7 v3.0 messaging in breast screening |
| A thorough understanding of the appropriate techniques | Qualitative and quantitative research techniques were employed for neural net testing and GP knowledge sharing |
| Ability to make critical use of published work and source materials | Valuable references such as Richards et al., Bankhead et al., Kalra’s PhD project on EHR, Dwivedi’s project work on KM framework are to name a few, for a more complete list of source materials, refer to the reference list |
| Appreciation of the relationship of the special theme to the wider field of knowledge | • core concepts of knowledge  
• KM framework and models in healthcare can be easily applied in any domain |
| Worthy, in part, of publication | • seven conference proceedings  
• two journal papers were published  
• two more journal papers are work in progress  
• refer to the list of publications for more details |
| Setting down a major piece of new information | • applying KM in real-time healthcare projects  
• comparison of healthcare messaging standards  
• breast screening attendance prediction-based messaging are addressed for the first time |
| Continuing a previously original piece of work | Arochena’s pioneering empirical work on breast screening attendance prediction was continued to prototype level |
| Showing originality in testing somebody else’s idea | Arochena’s neural network ROC characteristics was tested on new dataset and extended to new episodes |
| Carrying out empirical work that | Empirical study on GP’s views, attitudes and |

All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org
Objectives accomplished

This section details the accomplished objectives. It compares them to the list (set in chapter 1) of intended objectives.

- As part of the research methodology, semi structured interviews were conducted (through open-ended questions) which assisted in identifying the appropriate research direction and focus. The BSP call/recall process was thoroughly investigated as was the viewpoint of GPs (as part of the interview-based pilot study); it was crucial to understand the opinions, beliefs and expectations of these “knowledge receivers” to employ effective interventions as a strategic tool for increasing screening attendance.

- The literature review investigated the suitability of Electronic Data Interchange For Administration, Commerce and Transport (EDIFACT) and Health Level 7 (HL7) standards. As no comprehensive comparison of these standards were discovered during the literature survey, the project incorporated a comparison table. A comprehensive report which included design methodologies and implementation details for both standards was also included in the project report. This report highlighted the appropriateness of HL7 version 3 for this knowledge exchange.

- The design and analysis of a questionnaire to evaluate the GPs’ role in reducing BSP non-attendance was accomplished. This questionnaire not only identified the GPs’ opinions, beliefs and attitudes for reducing non-attendance through intervention but also collated the required information (for the correct timeframe) to share the predicted knowledge. The most important finding of the questionnaire was that not all GPs were expecting to receive financial compensation for their efforts to address screening non-attendance. A sizeable number of GPs reported that they would initiate interventions for any form of appreciation.

- The project identified a messaging platform and designed a viable architecture for integrating the Java-based Attendance prediction by Artificial Intelligence for Breast Screening (JAABS) algorithm. This algorithm was not only tested with a dummy dataset but also used real-time data collected from 1995 to 2008. Complete automation was incorporated within the proposed protocol right from the stage of raw data processing to the final prediction list. As a continuous improvement initiative, a new predictor attribute was also added to the existing predictor set. In addition, the quantitative analysis validated (through triangulation methodology) JAABS and AI-ATT to Logistic Regression.

Table 17 - Project work's originality summary

<table>
<thead>
<tr>
<th>has not been done before</th>
<th>beliefs on screening attendance prediction based interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being cross disciplinary and using different methods</td>
<td>This project work is a true cross-disciplinary work spanning areas from KM, AI, health informatics, healthcare, ICT</td>
</tr>
<tr>
<td>Adding knowledge in a way not done before</td>
<td>This project is a seminal work on prediction-based screening attendance improvement, using available technologies through the appreciation of humanistic nature of knowledge and its management</td>
</tr>
</tbody>
</table>
A complete Knowledge Management-based protocol (BSAMP) was developed ready for knowledge sharing with healthcare stakeholders via the NHS’ National Care Records Services (NCRS) infrastructure. The BSAMP protocol included the JAABS algorithm as the knowledge creation component. This project work was not limited to the design of the architecture but also created real-time message templates that can be instantiated to deliver the predicted knowledge. These templates covered both the pre and post-screening knowledge sharing stages. These templates were created using the HL7 version 3 message developing tools suggested by HL7 UK and NHS CfH.

Project deliverables accomplishment list

- evaluation report on the suitability of HL7 v3.0 standard
- comparison of HL7 v3 vs EDIFACT
- BSAMP message design and instantiation
- architecture for proposed knowledge sharing through messaging
- implementation details of JAABS algorithm
- design for prototype (JAABS algorithm)
- prototype execution and testing
- questionnaire survey results and recommendations
- reports, publications and presentations (refer to list of publications)

Adopting the proposed protocol

<table>
<thead>
<tr>
<th>Screening Year</th>
<th>Number of women invited</th>
<th>Screening non-attenders</th>
<th>Probable number of cancers not detected</th>
<th>Prediction of non-attenders using JAABS @42% NPV rate (till 6th episode)</th>
<th>Approximate number of women’s life saved, even if 25% of the interventions were effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996-97</td>
<td>1,558,995</td>
<td>388,190</td>
<td>1,776</td>
<td>728</td>
<td>182</td>
</tr>
<tr>
<td>1997-98</td>
<td>1,668,476</td>
<td>408,777</td>
<td>2,635</td>
<td>1,080</td>
<td>270</td>
</tr>
<tr>
<td>1998-99</td>
<td>1,669,727</td>
<td>409,084</td>
<td>2,580</td>
<td>1,058</td>
<td>264</td>
</tr>
<tr>
<td>1999-00</td>
<td>1,811,541</td>
<td>424,951</td>
<td>3,110</td>
<td>1,275</td>
<td>318</td>
</tr>
<tr>
<td>2000-01</td>
<td>1,815,610</td>
<td>453,903</td>
<td>3,286</td>
<td>1,347</td>
<td>336</td>
</tr>
<tr>
<td>2001-02</td>
<td>1,752,526</td>
<td>429,369</td>
<td>3,237</td>
<td>1,327</td>
<td>331</td>
</tr>
<tr>
<td>2002-03</td>
<td>1,873,470</td>
<td>473,988</td>
<td>3,539</td>
<td>1,451</td>
<td>362</td>
</tr>
<tr>
<td>2003-04</td>
<td>1,998,989</td>
<td>495,750</td>
<td>4,284</td>
<td>1,756</td>
<td>439</td>
</tr>
</tbody>
</table>

Table 18 - Probable number of life saved by JAABS algorithm

The objective of this project was to identify the challenges which are being faced by the UK NHS’ national breast screening programme and find approaches to alleviate these impediments and eventually reduce mortality due to breast cancer. While conducting the
literature review, the static attendance was identified as a persistent challenge and addressing it through KM was the driving force behind this research work. Table 18 provides a hypothetical scenario with current data to highlight the screening non-attendance, probable number of cancers not detected and approximate number of lives that can eventually be saved through the proposed initiative. This table provides an insight into the algorithm’s efficacy in saving human lives. When the proposed JAABS algorithm is deployed as envisaged in this research, it can considerably decrease breast cancer mortality. Based on the JAABS algorithm’s negative prediction value (for the first to sixth episode) the number of non-attendees correctly predicted a priori to screening date would be at least 42 women for every 100 screening non-attendees. When such knowledge is shared with the GPs (with whom the women are registered with) can initiate interventions. Such interventions can educate the non-attending women and clarify their attitudes and beliefs (Bekker et al. 1999). The expected outcome is that the women commits to a positive informed decision, which would culminate in attending the screening appointment.

A 25% success in GP interventions will result in saving more than 350 women’s lives per year. Even if one women’s life can be saved by our approach, this approach can be deemed a success. The new bespoke software prototype, incorporating JAABS algorithm can be easily converted and integrated in to the NBSS software. The messaging part has not been prototyped, as it has been envisaged to utilise the NCRS architecture operating on SPINE and DTS services. These services are already being field-tested at cluster level and once the national network is established would form the backbone of all NHS’ future clinical messaging across the country including the proposed protocol (BSAMP). The current work has already designed HL7 v3.0 messages for delivering the knowledge package as part of the new protocol (BSAMP). This message design has been developed on the same modus operandi followed by NHS CfH for easy integration. The ready to use message would ensure rapid deployment of the BSAMP at the national level with little overheads and modifications. As part of the current work, a questionnaire-based survey conducted with the GPs has identified the need of additional resources and recognition for making such future endeavours a complete success. Any future BSAMP-based initiatives for addressing the screening attendance should undeniably approach the issues from the KM perspective for ensuring positive results in improving breast screening attendance.

Challenges

Any research at the doctoral level has to work on the periphery of the corresponding domain’s knowledge boundaries. This creates numerous challenges to operate with limited resources. The requirement of continuous innovation at the highest level is fundamental for research work. Such exercises will certainly enable in developing new ideas, strategies and models. This section describes the challenges encountered in this research and how these were dealt with and solved. A number of intriguing challenges were encountered and solved during the current research work. This research provided a unique opportunity to better understand the nuances of applying KM in healthcare. It assisted in realising that there are numerous activities which business domains expedite as part of their routine activities even without knowing their importance and its relationship to effective KM. The “people” concept in KM is paramount and is the driver for any successful KM project.
GPs’ perceptions on primary care and their apprehensions on how their care services delivery have been transformed into small business domains and their impacts on healthcare were highlighted. This had placed in a lot of undue pressure that can result in counter productivity. GPs are the backbone of health services and are in direct contact with the public and their efforts are to be valued. Due recognition, even simple appreciation, can be leveraged as great motivators. These fundamental factors are all KM driven concepts of people being the core of any knowledge related activities and services. During the execution of this research work, it was found that convincing the stakeholders with regards to the importance of KM in healthcare was challenging. The GPs were reluctant to spend their resources on non-remunerative services i.e. addressing screening attendance. Future screening initiatives on screening attendance have to address these issues at a core level; failing to do so will force such initiatives to fail from achieving their true potential.

The current study relied on cancer data related to screening women’s historical episodes and demographic details for training the JAABS algorithm. This necessitated ethical approval for the whole research. This was achieved first by applying and obtaining the consent from the university’s ethics committee. Following this, the NHS Local Research Ethics Committee (LREC) was approached for getting approval. Even though the approval process was lengthy, and time-consuming, the whole exercise gave an opportunity to better understand the ethical issues related to clinical and non-clinical research.

The LREC approval meeting not only assisted in explaining the current research’s impact on screening attendance to the committee members, but was seen as an opportunity for acquiring new perspective and out-of-the-box ideas from connoisseurs having diverse expertise. The WSCBS unit had insisted to keep the dataset which contained the women’s postcode within the screening premises to ensure privacy, data protection, and security. To address this issue the prototype’s data pre-processing module was made as an individual entity that can consume the flat file generated by the screening unit's system. This data pre-processing module processed the input data and the postal code was replaced with two attributes namely the ‘postannum’ value, the “Townsend” deprivation value. The screening distance was calculated and a new variable containing only the distance-based categorical value was generated. Finally, this module gave an in process file that can be ported to the laboratory environment for further processing such as training, validating and testing the JAABS algorithm.

Designing and implementing the JAABS algorithm within an Open Source environment was challenging. This can mostly be attributed to the fact that most of such earlier work in the 1990s was carried out in C and C++ environment. Availability of efficient commercial software platforms have reduced the need for development of bespoke software, such as the one which was created for JAABS algorithm. This has not only created a gap in the availability of experts in this area but also has not given opportunities for creating such software solutions at research and development level. Commercial software platforms have their own inherent advantages such as ready made, time tested applications, better maintenance, support and a large choice of ready made algorithms that allows a mix-and-match strategy for exploratory type of work. Simultaneously such platforms also come with weaknesses, such as not being flexible; users need special training, lack of automation and non-availability of capturing repetitive work with less or no human involvement.
The Java environment being relatively new has less support for neural networks and hence had few prior initiatives in the said domain. One major handicap of Java was the memory recovery. The automated garbage collection made it difficult to release unused memory. Neural networks require massive computing resources. This fundamental requirement puts undue load on memory allocation and management, on the operating system and hardware resources. This challenge was partly overcome by using basic data structures available in Java and partly by creating in-process flat files for speeding-up the processing to avoid massive data to be stored on the Random Access Memory (RAM). The next challenging aspect of this research was the messaging protocol. This research work clearly identified that any new clinical messaging has to be focussed on futuristic needs of healthcare. Hence, the BSAMP protocol was planned to be based on HL7 v3.0 standard. The challenges can be attributed to very little earlier work done on this version and all the work being currently accomplished within the NHS UK were within a private environment and not publicly accessible; this resulted in immense challenges.

Such challenges slowed down the process of designing the HL7 v3.0-based messages for the BSAMP. This was partly overcome by attending special training sessions organised by the HL7 UK and its affiliated organisations. Further, the HL7 parent organisation has provided free tools and guides for HL7 members and new message developers. In addition to this, the support provided by HL7 UK and the NHS CfH was commendable; their assistance greatly sped up the message design and implementation for the current work. The above section specifically summarised some of the challenges met and how they were resolved. In addition to this, continuous changes in the NHS and process level enhancements implementation within the breast screening domain necessitated frequent updating of the strategy. Many micro issues that needed careful attention were duly incorporated by fine tuning the research focus so that the research objectives were achieved.

**Future work**

The project work being the first of its kind was limited by resources; hence it had focussed on core issues and culminated into a series of steps for the future research activities. The following section enumerates the probable future research direction that can be pursued to increase the breast screening efficiency and further the knowledge boundary.

1. The current work had concentrated on just one care deliverer i.e. GPs, for making interventions. This research work can easily be extended to other care deliverers such as specialist doctors for diabetes, cardiology, dentistry and care deliverers such as optometrist, pharmacist, orthopaedist, and physiotherapist. They are just to name a few who can make multi-pronged interventions. Such interventions can add value by increasing the probability of non-attendees to make an informed decision, thereby increasing screening attendance.

2. The algorithm is modelled in Open Source environment which employs object oriented concepts. With little modifications, this can easily be extended to predict other screening programmes actively pursued by the NHS National Screening Programmes (bowel cancer, cervical cancer, diabetic retinopathy,
Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

colon cancer, and prostate cancers). This may need exploratory research to identify predictor variables for their respective domains.

3. An immediate extension to the current work will be to identify a region and link the prediction results using the message design template for messaging through a web service-based email service to send messages to the GPs and update their database. A randomised controlled trial on at least three groups of GPs, one with the electronic messaging-based database updating, one group with the traditional paper-based prediction results transfer and one not making any prediction-based intervention. This study will validate the effectiveness with the electronic transfer of prediction knowledge when compared to traditional paper-based knowledge transfer.

4. A probable way to increase the efficiency of the JAABS algorithm is to identify more predictors. Ethnicity, educational background, women’s economical dependency, other healthcare related attendance etc can be included as predictors. This might need fresh data collection and collation for the screening women population. In addition new prediction models such as Support Vector Machines, genetic algorithm-based prediction techniques can be used as enhancements to the neural network. This kind of hybrid techniques will improve the current JAABS algorithm’s Negative Predictive Values (NPV).

Concluding observations

The current project has not only achieved all the planned objectives but also improved upon them and added value by developing new concepts. The project work additionally confirms that a research methodology can be an effective combination of both qualitative and quantitative research paradigms. A hybrid research methodology, which incorporates both paradigms, should also utilise triangulation as validating mechanism to achieve more credibility. This report sets out to highlight the conceptual developments that were part of the project and also clearly recognizes the distinguishing features with the earlier work (Arochena, 2003). In particular, the importance of the health informatics and KM components (which were not part of the earlier work) were added as new concepts to the existing knowledge domain.

Even though the JAABS algorithm was an adaptation of the earlier work (AI-ATT), this project focused on increasing the predicting efficiency and proved that such an algorithm can be incorporated within an open source framework with complete automation. The algorithm proposed in this project work is only a first step towards an automated prediction mechanism that would be used in a day-to-day functioning of the Warwickshire, Solihull and Coventry Breast Screening Unit. The ultimate goal is to extend the utilisation of the prediction technique to all the screening units in UK. An important improvement that has been implemented in the JAABS algorithm is to find more predictors for increasing its efficacy. One such indicator can be the women’s ethnicity. This attribute is difficult to collect for all the screening women population, but new research conducted in identifying the ethnicity by
the women’s name and other direct collation of information from the GPs and PCTs are also viable alternatives (Szczepura 2005).

The research was also the first of its kind, enumerating simple knowledge-based models and their implementation in preventive care. More than thirteen years of screening data was utilised to train the algorithm for the first time. The current work also proved that complex algorithms such as JAABS can be implemented in Open Source-based technologies. The research provides a fertile ground for fostering new research to continuously improve care services, specifically cancer surveillance. Clinical messaging is driven by the explosion of information and communication technologies. Electronic health records are fundamental to provide the right care, at the right place, at the right time and to the right person, irrespective of time and place where the care delivery is required. Even though HL7 v3.0 is relatively new and is not yet fully matured, further exposure and clinical implementation at an international level will provide enough opportunities for the standard to evolve and stabilise.

As part of this project, a qualitative analysis on health informatics and, in particular, electronic patient records was seen as an added deliverable. The discussion focussed on issues that are to be properly identified and resolved for effective electronic patient record. Special mention of academia’s involvement in developing new talents and the need for knowledge sharing at international level on health informatics issues were the focus of this work. All these issues have to be actively pursued for making global electronic patient record a reality. This project had revealed the effective use of primary care deliverers (GPs) for initiating the intervention for increasing uptake. A suggestion to extend this messaging framework to all the tertiary healthcare deliverers as interveners has to be validated and tested. The intricacies involving a whole gamut of healthcare deliverers may throw up fresh challenges and this can be tackled only by dedicated and well directed research in the said field of study.

In an earlier section, it was highlighted how primary care providers can be crucial in making preventive care a total success or failure. From the KM perspective, GPs are to be viewed as knowledge workers who are strategically positioned to make a huge impact on the population, thereby increasing the efficacy of such preventive healthcare services. Even inexpensive approaches such as appropriate recognition can yield positive results. One such approach of employing recognition (in the form of public appreciation of GPs efforts - specific to the value added services provided - such as trophies, commendations and awards) were proposed by the GPs. This work also proved that clinical messaging can be employed for efficient knowledge sharing. The research work covered in this project has indisputably underlined the importance of KM and seamless knowledge transfer across clinical domains. It further described how it can be successfully employed at empirical research. In addition, this work draws attention to the humanistic values that are to be addressed for a successful knowledge based research work. It further described how people (GPs), process (BSAMP) and technology (messaging, HL7, SOA, AI, XML) can work in unison on a healthcare KM project.

Finally, the research had proposed the use of HL7 v3 based messages for EDI and knowledge exchange. The HL7 v3 messaging has to be adapted for all future screening unit message exchanges. Further projects in healthcare domain have to appreciate the leveraging
abilities gained through appropriate use of KM. It should be mandatory for all research projects related to healthcare domain to focus on KM and its technologies to take healthcare delivery into the twenty second century.
References


Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)


Asuman, D., Gokce, B.L., Yildiray, K., Seda, U., Sam, H. et al. (2006) 'Exploiting ebXML registry semantic constructs for handling archetype metadata in healthcare informatics.', Int. Journal of metadata, semantics and ontologies 1, (1) 21-36


Bekker, H., Morrison, L. and Marteau, T.M. (1999) 'Breast Screening: GPs’ beliefs, attitudes and practices.' Family Practice 16, 60-65


Bell, T.S., Branstool, L.K., Newcombe, R.G.and Barton, G.R. (1999) 'Interventions to improve uptake of breast screening in inner city cardiff general practices with ethnic minority lists.' Ethnicity & Health 4, (4) 277-84

All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org


Bryant, H. (1996) 'How should we interpret noncompliance with screening mammography?' *Canadian Medical Association Journal* 154, 1353-1355


Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)


Fox, S., Murata, P. and Stein, J. (1991) 'The impact of physician compliance on screening mammography for older women.' Arch. Internal Medicine 151, 50-56


Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)


All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org


McDonald, C.J., (1997) 'The barriers to electronic medical record systems and how to overcome them.' *Journal of the American Medical Informatics Association* 4, (3) 213-21


Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)

All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org


Patnick, J. (2004) 'NHS breast screening: the progression from one to two views, Journal of Medical Screening 11, (2) 55-56


Pemberton, J.D., Stonehouse, G.H. and Yarrow, D.J. (2001) 'Benchmarking and the role of organizational learning in developing competitive advantage.' Knowledge and Process Management 8, (2) 123-35

Pemberton, J.M. (1998) 'Knowledge management (KM) and the epistemic tradition.' ARMA Records Management Quarterly 32, (3) 56-62


Prasad, B. (2001) 'Total value management- a knowledge management concept for integrating TQM into concurrent product and process development.' Knowledge and Process Management 8, (2) 105-22


Prusak, L. (2001) 'Where did knowledge management come from?' IBM Systems Journal 40, (4) 1002-7

Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)


All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org


Szczepura A (2005) *South Asian population less likely to attend cancer screening than non-asian.* [online] Available from <http://www2.warwick.ac.uk/ newsandevents/pressreleases/south_asian_population/>


Implementation of a breast cancer screening prediction algorithm: a knowledge management approach
A PhD project funded by the NHS Cancer Screening Programmes
Carried out by the Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University (www.coventry.ac.uk/biocore/)


Wilson, C. (1999) Coverage and uptake in the NHS breast screening program. MSc Project, Coventry: Coventry University


All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org
Appendix:

Research Papers Published


All enquiries relating to this report should be sent to Dr Rajeev K Bali, r.bali@ieee.org