Author(s): Imafidon, Chris; Kans, Aaron; Edoh, Aloysius
Title: Human organ re-representation using UML and CMAUT
Year of publication: 2009
Citation: Imafidon, C., Kans, A., Edoh, A. (2009) 'Human organ re-representation using UML and CMAUT' Proceedings of Advances in Computing and Technology, (AC&T) The School of Computing and Technology 4th Annual Conference, University of East London, pp.235-245
Link to published version:
http://www.uel.ac.uk/act/proceedings/documents/FinalProceedings.pdf

HUMAN ORGAN RE-REPRESENTATION USING UML AND CMAUT

Chris Imafidon, Aaron Kans & Aloysius Edoh
CITE; University of East London; University Way, London, E16 2RD
Chris5@uel.ac.uk; kans@uel.ac.uk; edoh@uel.ac.uk

Abstract. Clinical data was captured and stored data using natural language (NL) in order to describe the human organs, their attributes and behaviour (Olsen et, 1998). Although this was an accurate form of data representation it created information overload, space complexity, inconsistency and erroneous data. To address the issue of data inconsistency and standardisation, clinical coding such as UMLS was used while for clinical interoperability and data exchange between users, NL7 was introduced. A survey conducted by (de Keizer et, 2000a) revealed that these methods are inadequate for clinical data representation hence the data re-representation technique (Haimowitz et, 1988) was introduced and used for modelling CIS with Entity Relationship Diagram (ERD) and (FOL)(de Keizer et, 2000b). However this model does not address the issue of information overload and space complexity. Hence, this paper presents an alternative approach where UML is used to capture human organs, their attributes and relationships. A new framework with built in algorithm converts the multiple attributes modelled in the class diagram into mathematical formalisation using the CMAUT. The logical expression serves as input to the optimisation algorithm to determine the optimal amount of data that must be retrieved for primary healthcare investigation. To evaluate the framework, mathematical operations were performed which revealed that the space complexity when using the CMA re-representation technique is $\theta (n + 1)$ compared to $\theta (2n)$ for non- *CMA*. This means less space is needed when the CMA with AND connector is used but for substitutable organs with OR connector the space complexity for both CMA and non- CMA representations have the same exponential expansion of $\theta (2^n)$. A t-test conducted on the amount of data required for investigation before and after optimisation gave a p-value of 0.000 which means there is a significant different between the two data sets. For epidemiological analysis the output of the framework was benchmarked against the output of a web based heart risk calculator and the single sample t-test conducted gave a p-value of 0.686 meaning there is no difference between two outputs. Thus this framework with data re-representation occupies less space as compared to others and can be used to calculate the risk factor of a heart patient.

1. Introduction:

Clinical information system (CIS) is a subset of the medical informatics (MIS) that deals with the application of information technology for monitoring, diagnosing and analysing clinical cases for disease management and eradication. Clinical data was captured, stored and represented using natural language (NL) in word or text format. Although this method allows medics to represent clinical data accurately and in detail, it creates information overload, data inconsistency and space complexity (Olsen et, 1998). To solve these issues, different methodologies were introduced to represent clinical data, which includes clinical coding that address the data inconsistency and standardisation problems. Therefore, most CISs are now designed and built using digital image representation such as DiCOM, clinical coding or languages for example NeSH, SNOMED and ICD. These coding/languages describe clinical concepts

and their relationships however their limitations are they were design using strict hierarchies or semantic network structures and were developed for specific applications (de Keizer et, 2000a). For example ICD was developed by WHO for international health statistical analysis and SNOMED for classification of diseases, health care financial and performance analysis. Despite these advantages the representation of clinical data and implementation of live CIS are limited by the complexity of the knowledge content and the structure of the clinical data (Hoelzer el, 2003). To address these complexity new methodologies such as NL7 using object oriented techniques and UMLS using descriptive logic have introduced to create interoperability and standardisation in CIS (Friedman el, 2001). However, these methods do not adequately represent clinical data but instead create the information overload which is encountered during data retrieval and transmission as well as provide excessive clinical information for primary care disease investigation and management. This excessive data does not only overload the computer systems but also provide the medics with huge amount of clinical data to work through which is time consuming and lead to errors. Therefore there is the need to appropriately represent the data and create a framework to determine the optimal data required for investigation. The lack of optimisation mechanisms in CISs compels medics to retrieve the entire clinical data for diagnosis or prognosis instead of the relevant information needed thus creating information overload in the computing environment. Other issues associated with information overload in data intensive system include erroneous transmission of clinical data, mismatch or lost of patients' records, datacenter failures and performance degradation when integrated with others. These issues are experienced in NHS IT and Virginia Medical Centre USA projects. Work done on optimisation of CIS include 1 using advance query language and XML; 2. "best -of -breed" ; and 3.Entity Attribute Value enhanced ERD (Chen el, 2000). These methods do not optimise the entire CIS and do not address the fundamental issue of data representation.

According to (Haimowitz et, 1988) in developing information system the data structure must be similar to the domain been modelled this concept is called methods of data re-representation This approach, which is a computing problem, has been adapted and researched on by (de Keizer et, 2000b). They first analysed the ontology and structure of the various medical languages and highlighted their limitations (de Keizer et, 2000a). Then they proposed a new clinical data re-representation, which uses Entity Relationship Diagram (ERD) formalism to express clinical concepts and their relationship. To complement the conceptual model they used first order logics (FOL) as an expressive instrument and formal mathematical specification to avoid ambiguity (de Keizer et, 2000b). This approach increases the usability of CIS and confirms that the use of descriptive logics is not enough for modelling CIS (Werner el, 2003) and not suitable for optimisation and mathematical analysis.

This work adapted the (Haimowitz et, 1988) approach to create an optimisation framework for CIS which consists of two components: Firstly, a clinical data re-representation using UML and combinatorial multiple attribute (CMAUT): The class diagram captures the various organ combination in the human body using composition, association, and inheritance notations which are extended to include AND for representing complementary and OR for substitutable organs. The objects

and attributes in the classes are expressed in logical statement using combinatorial theory wherein the association between organs that have been affected by a disease are expressed with logical connectors. The expressions are rewritten in clinical data re-representation mathematical format which acts as the input for the optimisation algorithms: 2. The algorithm uses utility function and Raman/Grossman transformation table (Raman el, 1994) or [0, 1] matrix to convert the logical expressions into constraint matrix. The LP module in the framework is used to determine the optimal amount of data required for disease investigation and health care management. The objective function required for the optimisation process is formulated using the utility function. The utility function is calculated using the expected and actual recorded utility of the affected organs and thus allows the clinicians to express or select their preferences when using the algorithm to analyse diseases. The optimisation framework facilitates the partial retrieval of clinical data, reduce the information overload, space complexity that are associated with data transmission and improve the performance of integrated CISs
This paper starts with an introduction to CIS and it is followed in section 2 by a discussion on the use of UML and CMAUT in data representation. CMAUT theory is not applied in MIS (Sanderson el, 2006) because it is difficult but this work has used the concept to create a clinical data in a mathematical format for health care and mathematical analysis. In section 3, the operation of the framework is discussed. The results from the framework are analysed in section 4. Functional text was conducted to evaluate the framework conducted using cardiology case study to test the null hypothesis that clinical data when optimised using the CMAUT framework does not reduce the amount of data required for primary care investigation and thus does not reduce information overload. Secondly an epidemiological analysis was conducted to determine the percentage risk of heart disease attacking a patient. The null hypothesis used is that there is no difference between the output of the framework and values from the web based heart risk calculator, which uses Framingham predictive algorithm. The paper concludes with a summary of the key issues and future trends.

2. Data re-representation with UML and CMAUT:

The human organ re-representation uses the class diagram in UML as a conceptual model and CMAUT theory for expressing the formal specification to avoid ambiguity. CMAUT is a decision making tool used in econometrics but not medical application because of its complexity (Sanderson el, 2006). In CIS, class diagram is used to capture and represent the relevant classes required to handle patient-centred activities in the problem domain. This type of modelling does not focus on diseases as a concept nor does it capture the organs, their relationship and the multiple attributes. For instance, (Ziad El Balaa, 2008) attempted to capture the kidney and its attributes but the model is not comprehensive and therefore confirms the criticism that the NL7 does not incorporate or suggest a methodology for modelling disease, organs and their relationship in RIM (Taylor, 2006)
This work has developed a framework for optimizing CIS with emphasis on the diseases, patient organs, their attributes and association between the organs. The framework subsumes the patient as an object that is made up of six subsystems namely

respiratory, cardiovascular, neurological, coagulation, hepatic and renal as in SOFA algorithms used in Intensive Care Unit (ICU) (Ceriani el, 2003). SOFA is a decision making algorithm used in ICU for determining the malfunctions of patient organs. These subsystems complement each other in their operation to achieve the requisite homeostatic utility function is a human body. In this framework, organs that complement each other in their operation to achieve their predefined goals are called complementary organs and are linked with the AND connectors and their prescribed goals are known as utility function. Likewise organs that can substitute each other in their operation in order to achieve their utility function are known as substitutable organs and are represented with an OR connectors. Therefore the SOFA system is represented in the extended UML as: A class disease affects a patient class which is made up of 6 subsystems where the subsystem complements each other and connected with a AND logical connector. This conceptual model in which the association between organs are represented with combinatorial logical operators OR and AND while the attributes are captured with multiple attributes (MA) theory is called combinatorial multiple attributes (CMA). For instance a hypertension disease that affects the organs; heart, kidney and brain while complementing each other to ascertain that the correct amount of blood and pressure flows through the body to achieve homeostatic are known as complementary organs. Again, pair organs such as kidneys, liver, eyes are substitutable because when one is malfunctioning the other act as a substitute to achieve the utility function.

UML allows us to create a conceptual model of the organs and their relationship using OO approach and also capture their different attributes. This data representation can be converted in programming language using the formal methods Z or VDM approaches by (Kans el, 2003). This approach is unique because it: 1. Allows CIS to be optimised and determine the appropriate amount of data that can be mapped and retrieved for disease assessment. 2. Facilitate clinical data to be analysed using mathematical based algorithm or the epidemiological mechanisms such as in Framingham, QRISK and ASSIGN methodology for assessing the risk of heart related disease; as well as expand them by including newly discovered attribute because of its multiple attribute nature (Ceriani el, 2003).

2.1. Application of UML and CMAUT:

The domain scenarios used to illustrate the application of UML and CMAT in CIS are: Kidney and Heart related diseases: Firstly kidney related diseases involve the two kidneys working together to perform the function of regulating the flow and extraction of liquid in the human body. However, each kidney can replace or substitute the other during their operation. According to (Guyton el, 2006), kidney related diseases are classified into two namely acute renal failures and chronic renal failure. This work focuses on Interarenal acute renal failure, which is the result of abnormal behaviour of the kidney themselves. Inferring from the above, the kidney is a super class with two subclasses which are left and right kidneys which are linked with the OR connector. The subclasses inheritance the characteristics of the superclass as in figure 1 below.

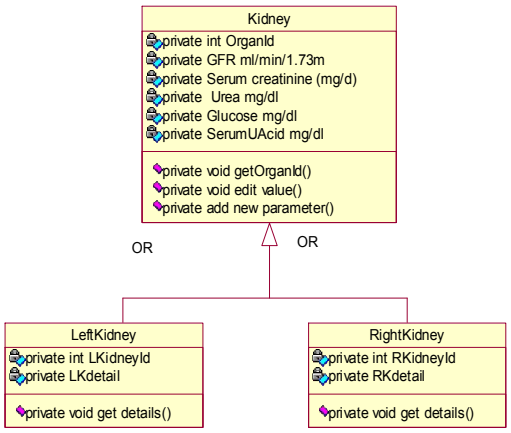

Figure 1; Clinical data re-representation of Kidney using UML

From Figure 1 the two kidneys can replace each others and have many attributes, hence are expressed in CMA using the OR connectors as

$X_2 \equiv [(K_1 \vee K_2), P_1, S_1, P_2, S_2]$ (1)

In (1) K_1 K_2 are the organs and P_1, S_1, P_2, S_2 are the attributes. The generic expression for this substitutable organs with multiple attributes is written as $[(C_1 \vee C_2 \ldots\ldots \vee C_n), P_1, S_1 \ldots Pn, Sn]$. This is further discussed in section 3 below

Scenario 2: heart related diseases include heart failure, hypotension, hypertension and angina: this project focuses on hypertension and hypotension. A hypertensive case is described as hypotension *(G)* is caused by high rate of pumping of the heart *(H)* which creates excessive blood pressure on the walls of the arteries (*A*) and sends appropriate signals to the brain (B) to regulate the flow of fluid in the kidneys (*K*). Hence hypotension and hypertension diseases affect three primary organs namely the heart, kidney and component in the brain (aka Antidiuretic hormone ADH) which complements each other in their operation. Figure 2 below depicts a class diagram with the three main organs and their attributes. The association between them is represented with AND operator. In this model the combinatorial organs with multi-attributes is expressed using logic connector AND in CMA as:

$X_1: [(H_1 \wedge K_2 \wedge B_3), P_1, S_1, P_2, S_2, P_3, S_3]$ (2)

In expression (2) the disease X_1 affects the body parts H_1, K_2 and B_3 where H_1 exhibits the attributes P_1, and S_1, while organ K_2 has attributes P_2 and S_2. etc. The generic expression for combinatorial clinical organs with multiple attributes using *AND* is $[(C_1 \wedge C_2 \ldots\ldots\ldots \wedge C_n), P_1, S_1 \ldots P_n, S_n]$. The logical expressions in (1) and (2) which serve as the input to the optimization framework are discussed in section 3 below

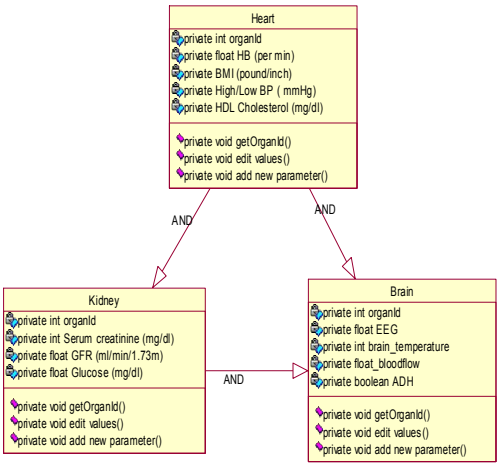

Figure 2; Re-representation of heart diseases with UML

3. Optimization Framework:

The data re-representation mechanism and algorithm in the framework works as follows: First the disease is defined as the organs or components in the human body whose attributes are performing abnormally or not to the specified norm (baseline). The relations between the organs in the disease domain are described using CMA that depicts the relationship between the various human organ affected and expressed in

multiple attributes as indicated in UML Figures 1 and 2 above. The multiple attributes in the expression (1) and (2) forms the basics for calculating the utility unit (U) of each organ using the formula (1.3). The weight assigned to each attribute is suggested by the medics and incorporated into the expression U = sum $[w_i\ f(s_i)]$ According to [54] in analysing heart related disease, heart beat and cholesterol must be given higher weighting but this depends on the medics who allocates the appropriate weight to the attribute. This is a clinical issue that requires further research. To convert the CMA expressions from the UML into CMAUT mathematical format which will serve as input for the optimisation framework the procedure below is used:

1. Write the relations to be optimized in logic expression;
2. Group the attributes and calculate the utility function using U = sum $[w_i\ f(s_i)]$
3. Use symbolic manipulation, to convert the logical expressions into conjunctive normal form (CNF);
4. Translate the logic expressions in CNF into linear mixed–integer variables (aka constraints) using Raman's Transformation table or [0,1] matrix
5. Establish the objective function to be maximize or minimize
6. Use the LP algorithm in the framework to optimise the objective function
7. Convert the evaluated value after the optimisation process to percentage
8. Map the optimal X values from the optimisation process with the attributes

To illustrate the above operations the logical expression for complementary organs are *G1* and *G2* with attributes as *P1* and P2 are first written as *(G1 G2, P1, P2)* while substitutable organs which is *G1* or *G2* with the attribute *P1* and *P2* are expressed as *(G1 G2, P1, P2);* Thus a disease *X1* that affect three organs becomes *X1: [(G1 ∧ G2 ∧ G3), P1, D1, P2, D2, P3, D3,]*. To convert these attributes into utility function (U) use the procedure below

$$U = \sum [w_i\ f(s)]\ f(\mathrm{s}) = \frac{Pi\text{-}Po}{Po} \quad (3)$$

The P_o is the expected pressure and the patient's measured pressure is P_i therefore the utility unit U_1 of the organ C_1 is shown in the (4) below as:

$$U_1 = \sum (w_{p1}\ f(sp_1) + w_{s1}\ f(sd_1) + w_q\ f(sq_1)) \quad (4)$$

Therefore disease X_1 is expressed in CMAUT format as X_1: $[(G_1 \wedge G_2 \wedge G_3)\ U_1, U_2, U_3]$ From (3) it is subsumed that the proposed algorithm takes into consideration the relationship between expected utility unit *Po* of the organ and the actual utility *Pi* measured at any time (t). The expected utility is the measure of the normal performance of an organ with respect to (w.r.t) the person's age, sex, height, weight while the expected attribute values are from medical literatures (Guyton el, 2006). The baseline parameter *Po* used for calculating the U_i was from (Guyton el, 2006) where the heart is working normally the systolic and diastolic pressures are: For up to 20years - 140/90mmHg; 20 to 50 years - 160/95mmHg and 50 to 75 years upwards - 170/105 mmHg. (In UK the generic standard blood pressure is 120/80). The heart rate or beat is between 70 to 80 times per minute. Similarly, the volume of filtrate formed by the two kidneys each minute is glomerular filtration rate (GFR) is about 125ml/min (i.e. 180 litres) a day. The renal blood flow is maintained at constant diastolic pressures of 80 to 200 mmHg. Again since the framework is benchmarked against Framingham algorithm the ADH in the brain

was not measured but specified as Boolean, similarly other kidneys parameters in Figure 1 and 2 were not considered.

3.1. Heart Disease Application:

To demonstrate the operation of the data re-representation mechanism in the framework the following example is used: This heart disease scenario in section 2.1 is formulated as: *(G)* "is caused by" high rate of pumping of the *(H)* "sending" excessive high pressure to *(A)* and "sending signal to" *(B)* to regulate the flow of fluid in the kidneys *(K)* These are complementary organs because they assist each other in performing their duties. Hence the combinatorial components with multi-attribute (CCMA) expression using logic connector AND is written as:

G_1: $[(H_1 \wedge K_2 \wedge B_3 \wedge A_4), P_1, S_1, P_2, S_2, P_3 S_3, P_4 S_4]$ (5)

G_1: $[(H_1 \wedge K_2 \wedge B_3 \wedge A_4), U_1, U_2, U_3\ U_4]$ (6)

In expression (5) the disease hypertensive G_1 affects the body parts H_1, K_2, B_3 and A_4, which exhibits the attributes P_i, and S_i, that are converted into utility units U_i in expression (6)

3.2.Determination of the optimal data using LP technique:

For the purpose of verification the results from the framework are compared with existing web based heart risk calculator that uses Framingham algorithm. Six web based heart risk calculators were analysed and their outputs found to be almost the same as confirmed in the survey conducted by (Chuang el, 2000). The heart risk calculator (Coronary, 2007) was selected because it is flexible and have the same attributes discussed. Five attribute values namely X_1, X_2, X_3, X_4 and X_5 that affect the combinatorial organs patient in hypertensive condition *(G)* were selected see Figure 2 and Table 2. The [0, 1] matrix was used where 1 represents attribute value measured and 0 indicates the attribute was not measured. The objective function to be optimize is

$$Z = \sum_{i}^{n} (U_1 X_1 + U_2 X_2 + .. + U_n X_n) \quad (7)$$

Table 1: Shows attributes values for organs

Organ Items	X_1	X_2	X_3	X_4	X_5
HB	1	0	0	0	0
BMI	0	1	0	0	0
BP 1	0	0	1	0	0
BP 2	0	0	0	1	0
HDL	0	0	0	0	1

In Table 2 the attribute values measured and recorded were converted into utility unit using the formula (3) and resultant utility values for the attributes are: $U_1 = 0.57$, $U_2 = -3.51$, $U_3 = 10$, $U_4 = 8.82$ and $U_5 = 6.25$. The objective function represents the utility unit offered by each organ (U_i) and the organ is represented as (X_i). The goal is to find the attribute in the combinatorial organs that has an overall utility unit that maximizes the utility value and to be retrieved for primary healthcare investigation. First, the above problem is presented as a LP with an objective function and a set of constraints as follows: Max or Min: $U_1 X_1 + U_2 X_2 + U_3 X_3 \ldots\ldots U_n X_n$, subject to the constraint matrix from Table 2.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

The objective function is rewritten with the values above as 0.57 X_1 -3.51X_2 + *10* X_3 + 8.82 X_4 + 6.25 X_5 and the constraint matrix shown in (8). The LP optimisation toolbox in MATLAB is used to find the optimal attribute valuation and the maximum value. The results after optimisation are optimal integer values $X_1 = 0$, $X_2 = 18.5$, $X_3 = 0$, $X_4 = 0$ and $X_5 = 40$. The solution indicates that the maximum value is `-314.935` and the optimal values $X_2 = 18.5$ and $X_5 = 40$ are the attributes which data are required for the investigation. Through mapping (see table 2) the data required are X_2 (i.e. *BMI)* and X_5 (i.e. *HDL*) while their risk factor is 3.14%. Again to measure the data size the attribute values for X_2 and X_5 were input into text file and measure on UNIX platform using the *mput* command.

To determine the space complexity the following mathematical operation was performed: The attributes in algebraic expression were converted into utility unit as in X_1: *[(C_1 ∧ C_2), U_1, U_2].* The expressions were transformed into conjunctive normal form (CNF) and to a set of inequalities (aka constraint matrix) using equational calculus and Raman's transformation table (Raman el, 1994) (Edoh, 2006) instead of [0,1] matrix used above. The optimisation is done by maximising or minimising the objective function $X = \sum_{i}^{n} Ui\ C_i$ subject to the generated constraint matrix. The output of the algorithm and the constraint matrix are optimised using LP technique to determine the MAX or MIN data required for clinical investigation. The results are shown in Figure 4 below where when the number of organs and attributes are increased the constraints also increases and likewise the computational space as in figure 4 CMA (aka CCMA) is y = *x + 1* and non-CMA *y = 2x (*aka CCSA)

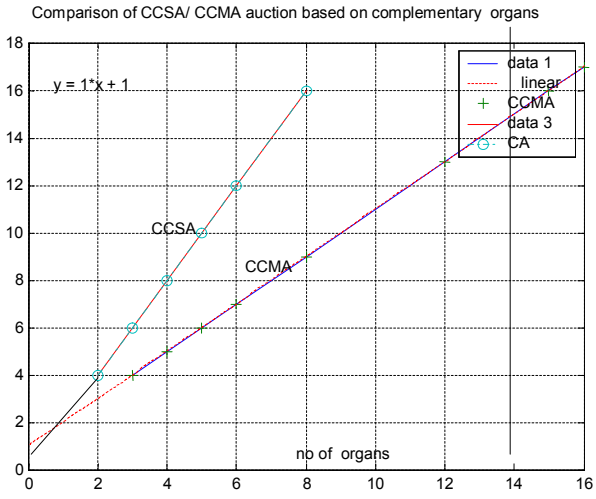

Figure 4 constraints for CCSA and CCMA using *AND* organs.

From the graph it is subsumed that disease that affect combinational organs and have multiple attribute (CCMA) required less memory space as compared to combinational organs with single attribute (CCSA) hence using CMA re-representation reduce load

4. Analysis and Discussion:

In this research, archived clinical records of 20 patients were randomly selected out of 100 from a cardiology research conducted at Withington Hospital [39] with UCL Health Informatics Centre, UK. The clinical data kept on UCL medical database is for research and subsumes that the selection of the attributes are accurate for clinical research because the measurement of

cardiology parameters +/- 3% tolerance range with under quality control regulations. In conducting the trial test only measurable attributes of the heart and kidney organs were used see Table 2 and Figure 5.

4.1 Functional analysis and test:

Methods used to validate the functionality of an algorithm or frameworks are time and space complexity (Kans el, 2003). The time complexity measures the performance of the framework in terms of the time required to execute the algorithm while space complexity estimates the resources required. Both methods use the big O -notation called Landau's (θ) symbol. a given execution time of an algorithm of *(n + 1) is* written mathematically in big O notation as *O (n) or θ (n)* which means the execution time is linear. Therefore for the graphs in Figure 4, which is the result of conducting a functionality test on the optimization framework the space complexity of *CMA* is *θ (n + 1) and θ (2n)* for non- *CMA* that is not optimised while using the AND connector. The research also indicates that the space complexity of CMA with *OR* as in kidneys Figure 1 and the non- CMA is the same exponential expansion $\theta (2^n)$

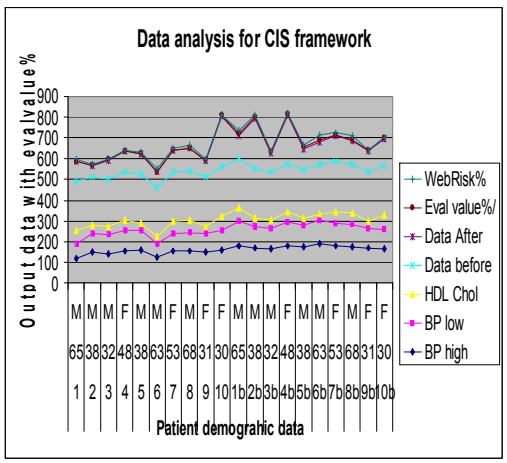

Figure 5 Variables used in test the framework

4.2. Statistical analysis:

The sample data for the trial was from 20 patients' records; t-test statistical analyses were conducted to determine the confidence interval (CI) and p-value. The CI for non-optimised patient data that was retrieved for primary care investigation before using the framework was 1.548 and after optimisation were 46.40. This means when clinical data is required for analysis the entire data is retrieved therefore the size is almost the same with minimal variation from the mean size of 237.3 MB. Using the optimisation framework, the standard deviation of the clinical data set was 46.40 with a mean value of 110.15 MB. This indicates that the mean data size required is low with great variation because the optimised data is specific to the patient's requirement instead of the entire data which overloads the network. The t-test analysis gave a p-value of 0.0000000866 which is less than 0.05 and means that the results are statistically significant. Hence the null hypothesis is rejected and the alternative hypothesis which states that the clinical information can be optimised using CMAUT framework to reduce the amount of data required for primary care investigation thus reducing information overload is accepted. This is confirmed in Figure 2 below and the second t-test conducted using the percentage difference between the data before and after optimisation gives a p-value of 0.000.

For epidemiological analysis the output of the framework was benchmarked against the web based heart risk calculator (Coronary, 2007) by recording the evaluation value of each patient from MATLAB see Table 2 below. The same data were entered into the heart risk calculator (Coronary, 2007) and the percentage heart risk calculated for each patient. The percentage difference between the expected heart risk and the observed

output value of the framework were calculated. Then a single sample t-test was conducted gave a p-value of 0.686 hence the null hypothesis which states that there is no difference between the output of the framework and web based heart risk calculator was accept. This means the framework can be used to determine the percentage risk of a patient been hypertensive however to extend the analysis to cover the entire population it recommended that Cohen's kappa agreement and clinical trial must be conduct which is beyond the scope of this project.

Table 2: Sample records of 10 patients

PID	A	S	Bh	Bl	Chd	Ev%
1	65	M	120	70	65	3.2
2	38	M	150	90	38	3.9
3	32	M	140	95	37	1.4
4	48	F	155	100	50	1.9
5	38	M	160	95	35	1.7
6	63	M	125	65	32	3.1
7	53	F	155	85	57	2.3
8	68	M	155	90	56	1.8
9	31	F	150	90	34	1.9
10	30	F	160	95	68	3.9

5. Conclusion:

In concluding this framework can be used to retrieve partial data for primary health care investigation and management. The p-value of 0.000 means medics can retrieve preferred relevant information and reduce the overload on the computer system. It will also allow health administrators and researchers to filter through the current data intensive clinical information system and select relevant data for their work. This is because the space complexity of CMA re-representation is $\theta (n + 1)$ as compared to $\theta (2n)$ for non- *CMA* For health care analyst, this approach will facilitate the creation of a mechanism for translating the current clinical coding such as READ, SNOMED into mathematical format that can be used for performance measurement and statistical analysis. Lastly, the single sample t-test gave a p-value of 0.686 indicating that there no difference between the framework and web heart risk calculator. Hence it can be used as epidemiological tool to determine the percentage risk of a patient been hypertensive.

6. Reference:

Olsen Kai A. [a], Sochats Kenneth M. [b] and Williams James G. "Full Text Searching and Information Overload" *The International Information & Library Review* Volume 30, Issue 2, June 1998, Pages 105-122

De Keizer N. Abu-Hanna F. A., Zwetsloot-Schonk J. H. M.; "Understanding Terminological Systems I: Terminology and Typology"; *Methods of Information in Medicine* 2000 39 1: 16-21

Haimowitz, L J; Patil, R S; Szolovits, P :"Representing medical knowledge in a terminological language is difficult": *12. Annual Symposium on Computer Applications in Medical Care*; Washington; DC (USA); 6-9 Nov. 1988. pp. 101-109. 1988

De Keizer N. Abu-Hanna F., A.; "Understanding Terminological Systems II: Experience with Conceptual and Formal Representation of Structure"; *In Methods of Information in Medicine* 2000 39 1: 22-29.

Hoelzer S, Schweiger RK, Dudeck J; "Knowledge representation in digital medical documents for efficient information access and retrieval": *Stud Health Technol Inform.* 2003;95:451-6

Friedman C, Liu H, Shagina L, Johnson S, Hripcsak G. "Evaluating the UMLS as a source of lexical knowledge for medical language processing". *Proc AMIA Symp*. 2001;:189-93. PMID: 11825178 [PubMed - indexed for MEDLINE]

Raman, R and Grossman, I. E. "Modelling and computational techniques for logic based integer programming", *Computers and Chemical Engineering*, vol 18, pp 563. (1994).

Edoh Aloysius, "Combinatorial Multi-attribute Auction (CMOA) Framework for E-Auction" *In the proc. of IEEE/SCSS conference on Advances in Computer, Information and Systems Sciences, and Engineering December 2005,* Springer published in 2006 ISBN-13: 978-1-4020-5260-6 pp 387 - 396.

Werner Ceustersa, Barry Smithb, Jim Flanagana "Ontology and Medical Terminology: Why Description Logics Are Not Enough" *Towards an Electronic Patient Record* (TEPR 2003), San Antonio 10-14 May 2003, Boston, MA: Medical Records Institute (CD-ROM publication)

Chen Roland S., MD, Prakash Nadkarni, "Exploring Performance Issues for a Clinical Database Organized Using an Entity-Attribute-Value Representation;" *J Am Med Inform Assoc.* 2000;7:475-487. DOI .© 2000 American Medical Informatics Association

Guyton Arthur C. and Hall John E. *Human Physiology and Mechanisms of Disease*, 6th Edition 2006 ISBN: 0721632998 http://medicalawards.elsevier.com/product.jsp?isbn=9780721632995

Taylor Paul. *From Patient Data to Medical Knowledge: Principles and Practice of Health Informatics* Blackwell Scientific 2006.

Kans Aaron, Charatan Quentin, *Formal software development from VDM to Java*, Palgrave Macmillan (9 Sept 2003) ISBN-10: 0333992814

Sanderson Colin, Gruen Reinhold, *Analytical Models for Decision Making Understanding Public Health*, Open University Press; Pap/Cdr edition (1 Mar 2006) ISBN-10: 0335218458

Chuang Jen-Hsiang, Kukafka R, P.H., Lussier Y A., R.A. Jenders J J. Cimino. "A Web-Based System for Prediction Of Coronary Heart Disease Risk Using The Framingham Algorithm" *In Proc AMIA Symp.* 2000; 983

Ziad El Balaa1 R, Traphöner; "Case-Based Decision Support and Experience management for ultrasonography " accessed on December 2008 http://ftp.informatik.rwth aachen.de/Publications/CEUR-WS/Vol-67/gwem2003-BT.pdf

Ceriani Roberto, Maurizio Mazzoni, Franco Bortone; Sara Gandini,Costantino Solinas, Giuseppe Susini, and Oberdan Parodi; "Application of the Sequential Organ Failure Assessment Score to Cardiac Surgical Patients"; *Chest.* 2003;123:1229-1239.)
American College of Chest Physicians

Coronary Heart Disease Risk Calculator; "On line clinical Calculator from the Framingham Heart Study"; http://www.intmed.mcw.edu/clincalc.html