International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 556 - 560, 2016, ISSN:- 2319–8656

A Study of Intrusion Detection System Methods in Computer Networks

Mohammad Hossein Karamzadeh
Islamic Azad University, Bushehr Branch,
Computer Group, Bushehr, Iran

Reza Sheibani
Islamic Azad University, Mashhad branch,
Computer Group, Mashhad, Iran

Abstract: Intrusion detection system (IDS) is an application system monitoring the network for malicious or intrusive activity. In these systems, malicious or intrusive activities intrusion can be detected by using information like port scanning and detecting unusual traffic, and then they can be reported to the network. Since intrusion detection systems do not involve predefined detection power and intrusion detection, they require being intelligent. In this case, systems have the capability of learning. They can analyze packages entering the network, and detect normal and unusual users. The common intelligent methods are neural networks, fuzzy logic, data mining techniques, and genetic algorithms. In this research, the purpose is to study various intelligent methods.

Keywords: Intrusion Detection, Normal Network, Genetic Algorithm, Computer Networks

1. INTRODUCTION

Public and private organization and institution increasingly use internet world wide network and the various services. In this world networks, millions of computers are connected to each other, and provide different services for millions of users. The important challenge of organizations is accessibility, control of users, internal and external, to network information and services. Users may endanger the security of network information. By information security, we mean integrity, confidentiality or availability of information. Intrusion refers to any set of actions that do not compromise one of these principles. In order to confront with systems intruders, intrusion detection systems are designed. These systems are located over a host or network, and detect intrusion on the basis of detecting misuses, unusual behavior or a combination of them [1]. Information security, intrusion and attack are defined. Then, intrusion detection systems are introduced. Also, performance, architecture and their techniques are presented. In addition, intrusion detection methods are introduced. Finally, two high-applicable methods of implementation are presented on the basis of fuzzy logic and neural networks.

2. INTRUSION DETECTION

Intrusion refers to actions that are not legally allowed, and endanger, three principles of information security involving confidentiality, integrity and availability. The intrusion to network is usually considered as an attack. The reason of these attacks can begin with a simple curiosity, and it continues to malicious and destructive objectives. In order to present, detect and stop the attacks, we must be able to detect time and the position of an intruder at first so that damages of organization information resources are minimized. Intrusion detection systems are responsible for detecting illegal misuses of the system or damages by both internal and external users. Intrusion detection systems are created as software and hardware, and each one has its own special advantages and disadvantages, speed and accuracy are advantages of hardware systems. Lack of security failure occurring by intruders is another advantage of such systems. Easy application of software, flexibility capability in software conditions and the difference of various operating systems dedicate more generality to software systems, and such systems are generally more appropriate options.

2.1 The history of intrusion detection systems

By increasing speed, efficiency, and connection of computers in 1970s, security systems are highly required. During 1977-1976, standard international organization held a meeting between governmental and inspection organs of EDP (Electronic Data Processing). The result was preparing a report in terms of security conditions, inspection and systems control. At this time, U.S.A ministry of power performed a research about inspection and security of computer systems because this country was concerned about security of its own systems. James P. Anderson was responsible for this mission. Anderson was the first person who presented an article about the necessity of automatic inspection of systems security. Anderson's report prepared in 1980 can be introduced as initial core of intrusion detection concepts. In this report, some mechanisms are introduced for systems security inspection. Also, it was shown how to control the system when a failure occurs.

During 1984-1986, Dorotty Denning and peter Neumann performed a research about the security of computer systems. The result was to create a real-time intrusion detection system performing as expert systems. This system was called IDES (Intrusion Detection Expert System). In this project, misuses detection was investigated. The idea proposed in this project was used as a base for many intrusion detection systems.

2.2 Intrusion Detection

- Audit analysis project

During 194-1985, a group began performing a project in Sytek with the command of America navy. The purpose of this project was presenting an automatic method to collect shell data for Unix operating system. Then, collected data are analyzed. In this project, separating desirable behavior from undesirable behavior is demonstrated [2].

- Discovery

Discovery is a system based on expert systems, and it is created to detect and prevent the problems in information bank of TRW Credit Company, this system is different from intrusion detection systems of that time. Unlike intrusion detection systems investigating operating system activities, this system investigated logs of information banks. The purpose of discovery is to prepare a report of unhallowed performances in information bank. In this project, statistical models are used to analyze data, and they are written in Cobol language [3].

- Haystuk

This project was performed by Haystack laboratory (1989-1991) and Tractor applied science (1987-1989) with the request of America Navy, The purpose of Haystack implementation was to provide an opportunity for security officers to detect misuses of SBLC (Standard Base Level Computer) computers of air force. These computers were mainframes 1100/60, and vintage operating system was performed in these computers. Processor motor of this system uses written to SQL and ANSIC language. This system can detect anomalies by using batchmode. In this case, information is continuously collected, and processed [4].

- MIDAS

MIDAS (Multics Intrusion Detection Alerting system) was implemented by NCSC (National Computer Security Center), and the purpose was to investigate Dockmaster systems (for which Hnoywell DPS 870 operating system is applied). In this system, information is collected and classified. Then, information of each class showing a connection and relation is compared with users' behavior. According to this comparison, they could detect false and unusual behaviors. MIDAS was implemented by LISP language. In this system, expert systems are used for processing [3].

- NADIR

NADIR was implemented by computer laboratory of Los Alamos, and it was used to investigate individuals performance on ICN network (Integrated Computing Network). ICN network is the main network of Los Alamos, and more than 9000 users use it. NADIR investigates the network by using collected information. In this system, statistical methods and expert systems are used for information processing [5].

- NSM

NSM (Network System Monitor) was implemented by California University. This system can be considered as the first intrusion detection system, and it uses information network as information resource. Previously, other intrusion detection systems performed their own tasks on the basis of information collected from operating system or programs' logs. Then, NSM performance and efficiency was used in most products.

3. GENERAL NERAL ARCHITECTURE OF INTRUSION DETETION SYSTEMS

An intrusion detection system generally involves the following parts.

- Information collection or sensor

This part is responsible for collecting information. For example, this part must create detect changes occurring in system file or network performance, and must collect required information.

- System Review

Each intrusion detection system should involve a part investigating the system itself in terms of its performance and efficiency so that accuracy and performance of the system can be assured.

- Information storage or information bank

Each intrusion detection system stores its own information in a place. This place can be a simple text or information bank.

- Control management

The user can create connection with intrusion detection system, presents necessary orders and commands.

- Analysis

This part of intrusion detection system is responsible for investigating collected information.

Architecture structure of IDS is observed in figure 1.

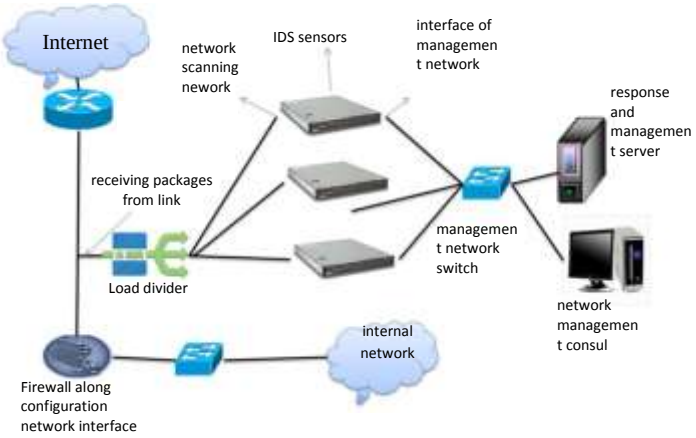


Figure 1: IDS architecture [5]

4. INTRUSION DETECTION

These techniques are used in misuses and anomaly detection systems. some of them are reviewed in this research.

4.1 Biological immune system

In this method, security in computer systems is proposed. The question is that "how a set of computers protect themselves?". In order to response it, similarities between biological immune system and compute immune system must be investigated. In a biological system, protection is performed by investigated. In a biological system, protection is performed by investigating finer factors like amino acids, proteins and etc. This can be true for intrusion detection systems. In this case, system calls can be considered as the first and detailed information resource. In this way, the order of system calls execution is statistically maintained for different users. If a program is executed, then execution order of system calls must accommodate with stored information. The alarm is presented if a difference is observed.

4.2 Genetic algorithm

One of intrusion detection methods is using genetic algorithm. According to genetic algorithm, intrusion detection process involves definition of a vector for information occurrences; that is, the related vector shows that occurrence is on intrusion or not.

At first, a hypothetical vector is considered, and its accuracy is investigated. Then, another hypothesis is proposed.

It is on the basis of test result, in previous hypothesis. Genetic algorithm involves two steps. The first step involves solution encoding as binary strings, while the second step involves finding a function to investigate binary string. GASSATA is a system performing the basis of it.

In GASSATA, system occurrences are classified on the basis of set of vectors. H (a vector for each string for occurrence) and n (the number of known attacks) are defined as follows. If it is equal to 1, than an attack occurs; otherwise, it is reserved if it is o. The function involves two parts. In the first part, danger probability of attack in a system is multiplied by the vector value. Then, its result is used to detect the error on the basis of second order function. In this way, false hypothesis are deleted. This step shows the difference between different attacks. The result of processing is to optimize the result analysis [7].

4.3 Statistical models in intrusion detection

Statistics is used in intrusion detection systems that are based on anomalies. Most of these systems have simple measurement tools, and determine attacks on the basis of changes in relation to a specified threshold limit. NID, of SRI apply expert statistical algorithm by using X2-like test of similarity measurement between short-term and long-term profiles. In our present statistical model, the algorithm similar to NIDES is used, but it has some differences. Therefore, some basic and main information about statistical algorithm of NIDES is introduced. The user profile is shown by the number of probability density function in IDES. S is considered as the sample space of random variable, and E1, E_2,…E_n events are considered in S sample space. suppose that

$$Q = N \times \sum_{i=1}^{k} \frac{(\acute{p}_i - p_i)^2}{p_i} \qquad (1)$$

Where pi is the probability of accruing Ei event. Also, imagine that pi is iterated continuously in specified time interval. N indicates the number of all events. In statistical algorithm of NIDES, X^2 like test is used to determine the similarity between the real and attack traffic.

When N is large, and E_1,E_2,…,E_n events care independent, Q is X^2 of (k-1) degree. since application programs cannot be immediately guaranteed, Q does not experimentally follow X^2.

NIDES solves this problem by using probability distribution function for Q updated daily in immediate operations.

Since we use neural networks to detect intrusion, we are not concerned about real Q distribution. The network traffic is not fixed, and it may be attacked in various times ranging from several seconds to some hours, so we require an algorithm for network traffic monitoring with different time window. According to observations, we use a window of static model layer (figure 1). Each model layer corresponds with time cut. Events occurring at present must be stored in layer buffer. 1. Stored events are compared with reference model of that layer. The result is transferred to neural network to detect the network position. When buffer of time event is full, it becomes empty, and then stored events are transferred to buffer of layer event 2. The similar processing is performed until reaching the highest level recursively. Events are easily removed after processing in the highest level. Similarity-meatering-algorithm whom we use is as follows:

$$Q = f(N).\left[\sum_{i=1}^{k} |\acute{p}_i - p_i| + \max_{i=1}^{k}(|\acute{p}_i - p_i|)\right] \qquad (2)$$

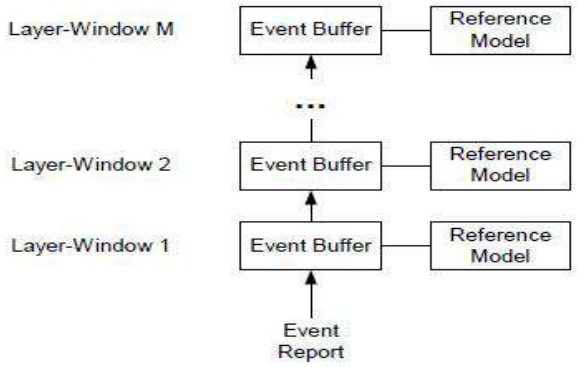


Figure 2: Statistical model [8]

f(N) is a function used to compute the number of all events occurring in a range of window. In addition to similarity measurements, we designed an algorithm to update the color of reference model. We consider P_{old} as the resource model before updating. P_{new} is considered as a reference model after updating, and P_{obs} is taken into account as the activity of a user in time window.

Updating formation of reference model is as follows:

$$\bar{p}_{new} = s \times a \times \bar{p}_{obs} + (1 - s \times a) \times \bar{p}_{old} \qquad (3)$$

In this formula, a is predefined adaptation rate, and s is the value produced by neural network. Suppose that output value of neural network is a continuous value between 1 and -1.In this case, -1 means absolute intrusion, and 1 is lack of absolute intrusion. Different values show related absolute levels. Computation function of S is as follow:

$$S = \begin{cases} t & if\ t \geq 0 \\ 0 & otherwise \end{cases} \qquad (4)$$

4.4 Neural networks

Neural networks are widely used as effective methods of patterns classification, but application programs are altered due to high volume of computations and long learning cycles.

BP neural networks are used by [4] and [7] to detect activity anomaly. In [2], we extend the example of hybrid neural network called hybrid backward perceptron network. This network is a combination of perceptron neural networks and small backward network. In order to understand neural network better, we tested five types of neural networks involving perceptron, RBF , FUZZY MAP, PBH and BP. Perceptron [9] of figure (3) is an sample of neural network used to classify linearly separable patterns. It only involves a neuron with setting threshold limit and connection. We use perceptron neural networks as a base to evaluate efficiency and performance of other neural networks.

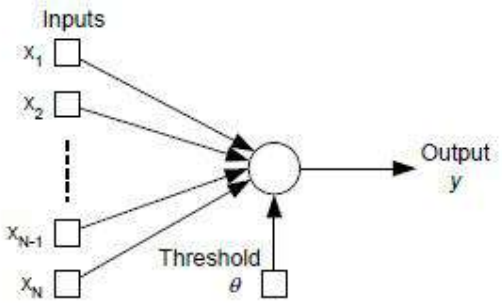


Figure (3) perceptron and architecture

Backward networks [9] or BP (figure 4) is a multi-layer forward network involving an input layer, on output layer and a hidden layer. BPs have higher production power, and they are used to solve some diverse and difficult problems. We tested BP network by some hidden neurons in range of 2-8.

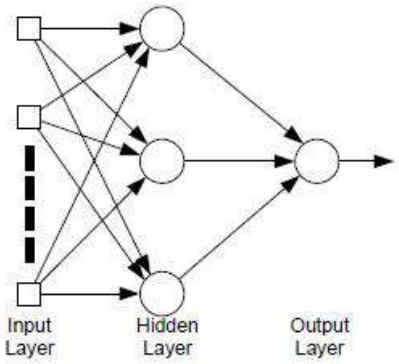


Figure (4): BP architecture [8]

Hybrid backward perceptron [6] or BPH (figure 5) is a combination of perceptron and small backward network. BPH networks have capability of linear and nonlinear detection and discovery, and they depend on input stimulus vectors and output values. We tested BPH neural networks in range of 1-8 of hidden neuron.

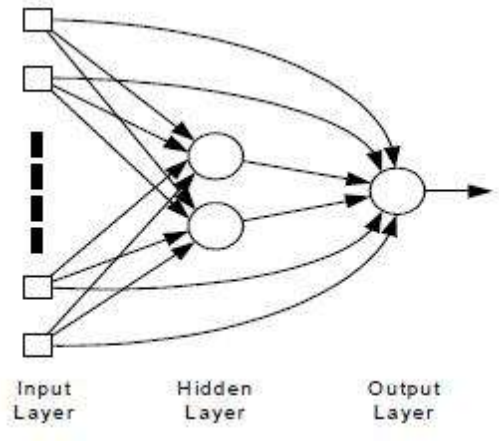


Figure (5): BPH architecture [8]

FUZZY ARTMAP [10] involves two fuzzy art networks: ART_a and ART_b. F2 layers are connected by subsystems introduced as match tracking system.

We used ARTMAP system [11]. Figure (6) is used to classify the problems. We tested ARTMAP neural networks with 1-8 neurons.

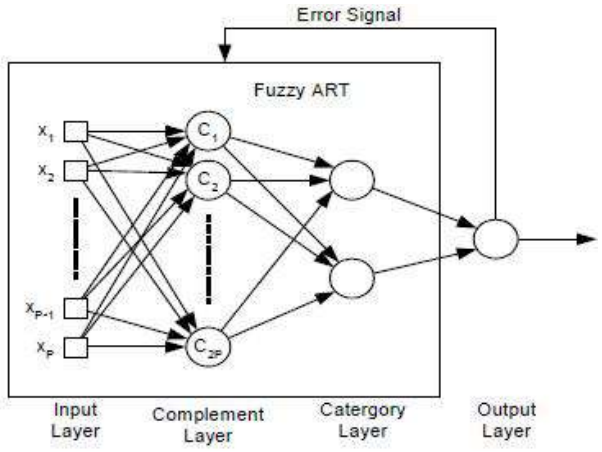


Figure (6): Fuzzy ARTMAP architecture

Racbal-basis network [9] or RBT (figure 7) involves three input layers. Input layer is constituted of resource nodes. The second layer is a hidden layer with enough large size, and presents different purposes of BP network. Output layer provides network replay to activation patterns applied to input layer. We tested RBT networks for hidden neurons in range of 2-8 [9].

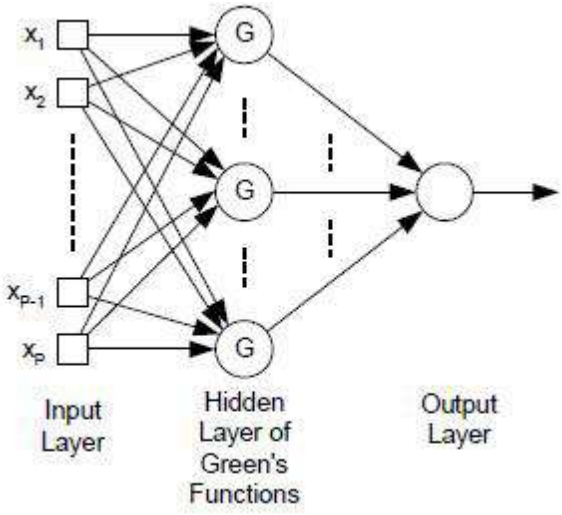


Figure (7): RBT architecture [8]

In our experiments, we used professional neural works of II/PLUSTM to generate all neural networks discussed earlier.

5. SUMMARY AND CONCLUSION

Intrusion detection systems are one of used tools to create security of computer networks. These systems can be classified according to two views. The first view is related to information resource, while the other one is the method of intrusion investigation. According to intrusion investigation method, two classes of systems can be considered such as the method of detecting misuses and anomalies. The research shows that IDS systems are very efficient to detect intrusion in network.

6. REFERENCES

[1] Crosbie, M., &Spafford, G. (1995, November).Applying genetic programming to intrusion detection.InWorking Notes for the AAAI Symposium on Genetic Programming.MIT, Cambridge, MA, USA: AAAI, pp. 1-8.

[2] Gong, R. H., Zulkernine, M., &Abolmaesumi, P. (2005, May). A software implementation of a genetic algorithm based approach to network intrusion detection. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks.SNPD/SAWN 2005. Sixth International Conference on, pp. 246-253.

[3] Hashemi, V. M., Muda, Z., &Yassin, W. (2013). Improving Intrusion Detection Using Genetic Algorithm.Information Technology Journal, 12(5), pp. 2167-2173.

[4] Li, W. (2004,May). Using genetic algorithm for network intrusion detection.InProceedings of the United States Department of Energy Cyber Security Group 2004 Training Conference,pp. 24-27.

[5] Lu, W., &Traore, I. (2004). Detecting new forms of network intrusion using genetic programming. Computational IntelligenceJournal, 20(3), pp. 475-494.

[6] Mukherjee, B., Heberlein, L. T., & Levitt, K. N. (1994). Network intrusion detection. Network IEEE, 8(3), pp. 26-41.

[7] Ojugo, A. A., Eboka, A. O., Okonta, O. E., Yoro, R. E., &Aghware, F. O. (2012). Genetic algorithm rule-based intrusion detection system (GAIDS).Journal of Emerging Trends in Computing and Information Sciences, 3(8), pp. 1182-1194.

[8] Shaveta, E., Bhandari, A., &Saluja, K. K. (2014,March).Applying Genetic Algorithm in Intrusion Detection System: A Comprehensive Review.In5th International Conference on Recent Trends in Information,Telecommunication and Computing(ITC 2014.India:ACEEE, pp. 102-112.

[9] Xia, T., Qu, G., Hariri, S., &Yousif, M. (2005, April). An efficient network intrusion detection methodbased on informationtheory and genetic algorithm. InPerformance, Computing, and Communications Conference, 2005 .IPCCC 2005. 24th IEEE International.IEEE, pp. 11-17.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 561 - 567, 2016, ISSN:- 2319–8656

Image Steganography using Fusion Based Advanced Encryption Algorithm and Embedding Techniques

Venkateshappa
Research Scholar
M.S Engineering College
Bengaluru-562110, Karnataka, India

Sunitha.P.H
Research Scholar
M.S Engineering College
Bengaluru-62110, Karnataka, India.

Gopala Krishna Murthy.CR
Electronics and Communication
KSSEM, Bengaluru

Abstract: In real time application of transmitting data through internet or web encryption plays a vital role. Data that is being transmitted will be encrypted in order to be protected from the hackers. When data is sent through internet it can be viewed by number of people. The data transmitted first moves to the local network and then to internet service provider, who will be able to view your message. After this the data travels through number of routers to reach the internet service provider of the recipient. During this process data can be accessed by numerous people. Hence, we need to encrypt the data or message that is being transmitted. Encryption can be done by several techniques and different algorithms. This project will embed three different techniques to increase the level of security provided to the data. Haar DWT and Average alpha blending are the techniques used along with that LSB encryption algorithm is used. In all existing techniques lift DWT was widely used where, it produces a negative co-efficient. Handling negative co-efficient will degrade the PSNR value of the restored image at receiver side. Hence, it is replaced by Haar DWT. Haar wavelet will be implemented on the input image which provides one level security and using LSB algorithm where output of Haar wavelet pixel values are altered by pixel values of cover image used. This step provides two level security. Average blending is done at the last level so that levels of security are increased again and also speed is increased area is minimized. In this project using all techniques mentioned above, security is increased and use of Haar wavelet will also increase PSNR value.

Keywords: DWT, Haar wavelet, LSB technique, average embedding technique

1. INTRODUCTION:

In present era security is important issue in communication and storage of information, images, audio and videos. This must be protected from mischiefs, it is applied to any vulnerable and precious resources. rseparation is done between the resources and risk. Security is important in many fields such as home security, computer security, banking security, information security, etc. Home security: Home security is applicable to all of us. The home consists of many things which must be secured. Example: the widows must be closed, doors must be locked properly. Security is important to avoid a robbery.

Computer security: Computer security is also called as cyber security, IT security. Security is applied to processing devices such as computers and Smartphone and also computer networks. Computer security includes five components they are: hardware, software, data, people and procedures by which information is restricted to the unauthorized access. It also includes physical and information security, physical security is to prevent the theft of equipment. Information security is to protect the data on the equipment.

Banking security: In a core banking system, there is Chance of encountering forged signature for transaction and in the net banking system, the password of customer may be hacked and miss-used. Thus, security is still a challenge in these applications. The main aim is to secure he customer information and to prevent the possible forgery of password hacking.

Information security: In the present situation the use of internet is being increased rapidly. The innovation of the technology has lead to increase in

the speed of transmission of data through communication channel which is easily exposed to the unauthorized person. Hence, there is a need for safeguarding individual's creation from the copyright. This can be done through a technique called Digital watermarking.

The Haar wavelet is a rescaled sequence "square–shaped" functions form a wavelet family or basis. The Haar sequence is now well-known as the first wavelet. Alfred Haar proposed Haar sequence in 1909, he used this function to give an example of an orthonormal system. Haar wavelet is also the simplest possible wavelet, it has orthonormal properties.

Alpha blending technique is used to blend or to insert the watermarked image. Cover image and watermarked image which is obtained after application of Haar DWT, they multiplied by a scaling factor and are added.

Data hiding in a grey scale image using LSB technique is simply replacing the LSB bits of the host image with the cover image. Each equivalent pixel of host image and each equivalent pixel of cover image are considered and then the LSB of the host is replaced by the LSB of the cover image it means that embedding the secret information in to the cover image. By making use of this technique the security enhances there by it is difficult for an unauthorized person to decrypt the secret information or image.

2. LITERATURE REVIEW

Steganography is masking of a file, message, image, or video. In this paper the Steganography is done on the boundary. This boundary based Steganography or steganalysis uses an auto-aggressive model [1].

Growing technology has made rapid increase in usage of internet. The advanced technology has lead to transmission of data through network which is easily exposed to the unauthorized person. Hence, there is a need for safeguarding individual's creation form copyright. This can be done through a technique called Digital Watermarking [2].

to enhance the security in a digital image captured by camera. As the image capturing has become passion for few people they wanted to publish their photography. But when there is mislead to tracking to theft the photograph would lead to great loss to that person. Hence the data must be secured carefully. Singular value decomposition (SVD) and DWT [3] is applied on the watermark of a RGB domain.

The Endeavour of this project is to detect the outcome of PSNR value by making use of implementing 5/3 2D lift DWT based watermarking technique [1], The expelled data during transmission and reception is preserved to its minimal by a technique called Alpha blending and resizing.

Compression of image and video now-a-days is unarguable. For different input patterns the multi level 2D DWT perform several computations for execution. The designs made used are 6 row-columns line-based and block-based [4].

The main criteria of the paper is to reduce the bandwidth of the image during transmission wavelet based technique such as JPEG2000 for image compression is the best method in compressing ratio[5].

The objective of this paper is to detect the power dissipation during data hiding using 2D-DWT using lifting technique. They have used a CoDel language which is a procedural language to order the statements implicitly represents the sequence of the activities[6][7].

In this paper they have made use of improved LSB based Steganography technique for image which gives better security. In the edges and sooth area of an image the secrete image or message is being hided in non-adjacent and random pixel locations. By making use of edge detection filter, the edges of the cover image is being detected and encrypted secrete image's edges will be replaced at the LSB of the cover image i.e., red, blue, green pixel components on randomly selected pixels on smooth area of the image [8][9][10][12].

The techniques used in this design are LSB, DCT and compression technique on row image. Steganography is nothing but hiding an image with in another image or video, text, etc. LSB technique [11] is to embed the payload bits in to the cover image which forms a stego image.

The paper explains that the image is hided in a frame of video they have made use of algorithm called frame decomposition technique. They have used three techniques that color map matrix of RGB image, LSB, CDT of RGB image and unique matrix

of RGB image. The output of CDT will have some limitation that is the image has maximum of 255 pixel values after the application of CDT maximum pixel value will be 255 only for a single image. As the Steganography is done on the image and a video by CDT [13].

Amid the embedding process we will see that the size of the watermark is smaller than the cover image. Here the edge size of both the watermark image and host image are made equal. Since the watermark embedded in this paper is recognizable in nature or unmistakable, it is inserted in the low frequency approximation component of the host image. After alpha blending technique we will acquire the watermarked image which comprises of the original watermark along with the original image [15].

1.1. Summary on literature review

From the above review we came to conclusion that the image Steganography is carried out using digital watermarking, LSB technique, DWT, Alpha blending using multipliers and adders, lifting DWT. The security of the image is enhanced and proved that the encryption technique does not degrade the image quality by making a comparison with the PSNR values between the input secret image and the decrypted secret image. The PSNR value of the image can still be increased by making use of Haar DWT and by making use of Average Alpha blending using multipliers.

2.2. Speed is increased and the area utilization will be less when compared with normal alpha blending technique using adders. Hence an area and security efficient architecture of image Steganography is proposed. Limitations of Existing Systems

By making use of 5/3 lift DWT negative co-efficient results so there by Haar DWT will eliminate the negative co-efficient. Edge details are taken in to consideration and the security is enhanced but most of the information will be present in LL band so concentrating mainly on the major part of the information which is LL band and the security will be enhanced. Alpha blending technique uses more bit storage area. Therefore using average embedding algorithm which reduces the bit storage values and results in reduction of memory.

.

3.DESIGN AND IMPLEMENTATION

3.1. General Block Diagram

Steganography of an image includes a series of steps. The general block diagram for increasing the security of communication is shown below in figure 1, where each and every block will be elaborated.

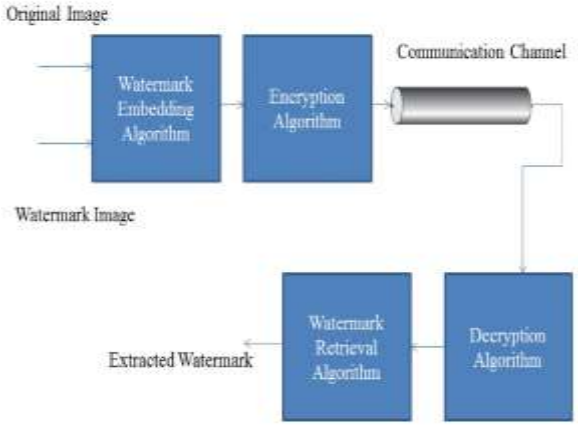


Figure 1 General block diagram

Initially the original image (secrete image) and watermarked image(cover image) is fed to the watermark embedding algorithm. Here, the embedding algorithm is LSB technique. The output of the embedding algorithm is then fed to encryption algorithm which is average embedding algorithm. The output of the encryption algorithm is passed through communication channel, it can internet or LAN, etc. At the other end decryption of all these algorithm is done by knowing which are techniques that is being used. Finally the image will be extracted.

3.2 . Proposed Model

The proposed model is as shown below in figure 2. DWT is nothing but the size of the image is compressed to increase the security Haar wavelet co-efficient are made used. To get the better performance additional technique is included that is LSB technique. The output of this LSB technique will be passed to average embedding algorithm which is also known as average embedding algorithm.

The next step is to verify the quality of images; the PSNR value is calculated between the watermarked images also known as encrypted image and the cover image. This forms a transmitter section encryption. To pass this encrypted image the size of the image should be as same as the original image, so applying inverse DWT and then passing it through the communication channel.

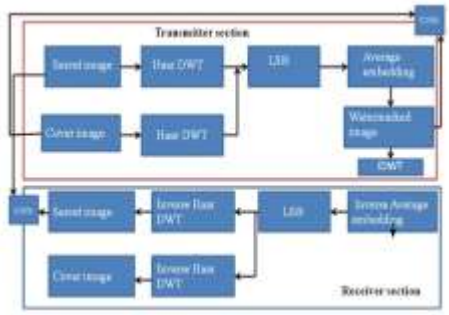


Figure 2 Proposed Model

The decryption must be done at the receiver section by applying inverse of all the techniques used, by doing this, the encrypted image will be retrieved. Finally, to know the percentage of security enhancement PSNR value will be calculated between the secrete image and retrieved secrete image and also between the cover image and retrieved cover image.

3.3 Haar DWT

DWT is nothing but Discrete Wavelet Transform, application of DWT to the image is to compress the size of it. The signals that are generated are translated into shifted and scale versions of the mother wavelet to generate DWT bands. Depending upon the wavelets chosen the security also increases. In this project Haar DWT is used to enhance the performance in terms of PSNR value and scaling the area.

Taking finger print as an example, decomposition of 2D Haar DWT is shown in figure 3.8.

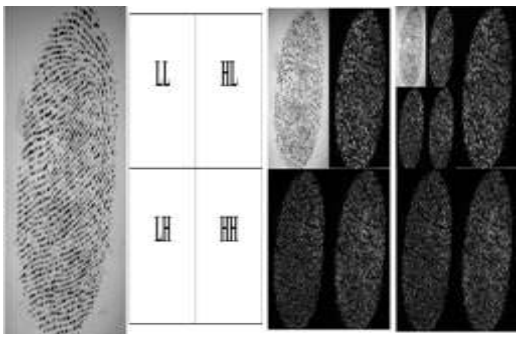


Figure 3 2D-DWT decomposition

The decomposition of fingerprint using DWT at two levels is shown in Figure 3.

The Haar wavelet has orthonormal properties i.e., orthogonal with unit vectors which is used as the mother wavelet and has simplest useful energy compression process. The Haar transformation of one dimensional input leads to two vector elements that is given by the equation 3.1.

$$(y(1), y(2)) = T(X(1), X(2)) \dots\dots\dots\dots\dots\dots 3.1$$

Where $T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is the Haar operator

y(1) and y(2) are sum and difference of x(1) and x(2) which produce low pass and high pass filtering respectively, it is scaled by $1/\sqrt{2}$ to preserve the energy.

The Haar operator T is an orthonormal matrix since its rows are orthogonal to each other that is their dot products are zero and have unit lengths, therefore $T^{-1}=T^{T}$. Hence we may recover x from y using equation 3.2.

$$(x(1), x(2)) = T^{T}(Y(1), Y(2)) \dots\dots\dots\dots\dots\dots 3.2$$

For 2D image, Let *x* be 2×2 matrix of an image, the transformation *y* is obtained by multiplying columns of x by T, and then the rows of the result by multiplying by T^{T} using equation 3.3.

$$y = T * x * T^{T} \dots\dots\dots\dots\dots\dots 3.3$$

The original values are recovered using equation 3.4

$$x = T^{T} * y * T \dots\dots\dots\dots\dots\dots 3.4$$

An Example of DWT,

If $x = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is the original matrix, then DWT is given in equation 3.5.

Then

$$y = \frac{1}{2}\begin{pmatrix} a+b+c+d & a-b+c-d \\ a+b-c-d & a-b-c+d \end{pmatrix} \dots\dots\dots\dots\dots\dots 3.5$$

The DWT bands correspond to the following filtering processes:

LL: a+b+c+d : Low pass filtering in horizontal as well as vertical direction.

HL: a−b+c−d : High pass filtering in horizontal direction and Low pass filtering in vertical direction.

LH: a+b−c−d : Low pass filtering in horizontal and High pass filtering in vertical.

HH: a−b−c+d : High pass filtering in both horizontal and vertical direction.

To use this transform to a complete image, the pixels are grouped into 2x2 blocks and transformations are obtained using equation 3.5 for each block. The 2 level DWT is applied on fingerprint image of size 256x256 to obtain 128x128 coefficients after first level and 64x64 coefficients after second level stage. The 64x64 LL sub-band coefficients are considered as DWT features. Haar DWT is as shown in figure 4.

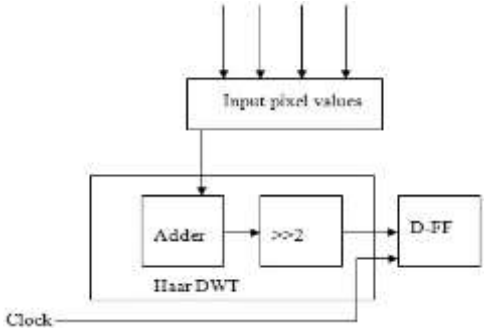


Figure 4 Block diagram of Haar DWT

3.4. LSB Technique

Least Significant Bit is encryption algorithm used to improve the security of communication through a communication channel. Taking two images that is one image is cover image and the other is a secrete image the operation will be performed. The idea behind this technique is to embed or hide the secrete image in a cover image. Embedding the image in an image is nothing but replacing the LSB of cover image with the MSB of secret image where LSB of secret image acts as a key. These bit values will be obtained from the previously obtained output that is from the Haar DWT output. The input to the LSB is the output of Haar DWT where the image size that is 256x256 of both cover and secret will be compressed to 128x128. At every clock cycle the input will be given to LSB technique from Haar DWT. At each and every clock cycle the pixel value gets encrypted. This operation continues till the complete image pixel value is covered that is 128x128.

For example: let the pixel value of cover image be 128 and 115 of secret image.

The binary equivalent of the pixel values are:

128:-10000000 and 115:-01110011

LSB algorithm:-10000000=cover image pixel

01110011=secret image pixel

10000111=embedded pixel

The decimal equivalent of embedded pixel value is 67 and 3 is the key for that pixel. Same process will be held to compute the entire pixels value to embed.

4. RESULTS AND DISCUSSION

A. Simulation Output for Haar DWT

The output of Haar DWT is shown in figure 5.

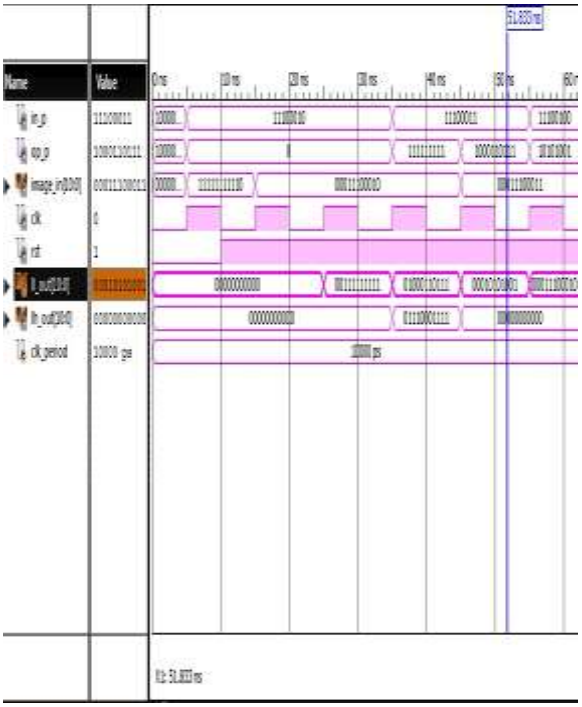


Figure 5 waveform of Haar DWT

4.1. Simulation Output of LSB Technique

The simulation output of LSB technique is show in figure 6.

Figure 6 waveform of LSB technique

4.2. Software implementation

The software implementation using system generator for Haar DWT is as shown in figure 7.

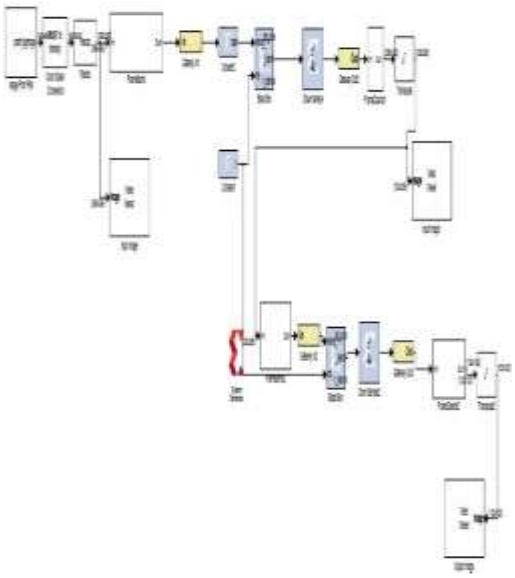


Figure 7 software implementation of Haar DWT

The software implementation using system generator for LSB is as shown in figure 8.

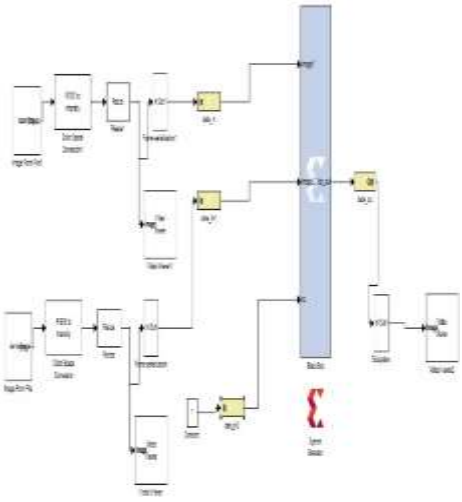


Figure 8 software implementation of LSB technique

4.3.Output of Haar DWT using system generator

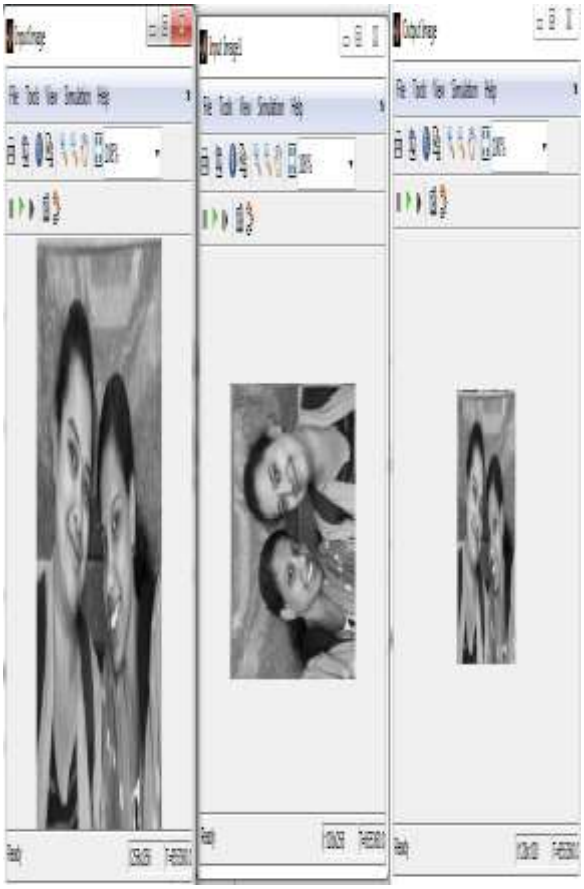


Figure 9 output image of Haar DWT

4.4. Output of LSB using system generator

Figure 10 output image of LSB technique

5. CONCLUSION AND FUTURE SCOPE

The proposed paper embedded different techniques for encryption of a image. Since application of only DWT and average blending was producing a single level security. Now, in this it gives a 2-level security.

In future, the same techniques will be applied for decrypting the secrete image by applying the inverse of the algorithms and techniques used for encryption. Further this can also be used for video file steganography.

6. REFERENCE

[1] B.Pushpalatha, Mrs. Shalini Shravan, "An Efficient 2D 5/3 Lift DWT Based Invisible Watermarking Technique", National Conference at KSSEM 2015.

[2] M. Jiang, X Wu, E. K. Wong, and N. Meinon, "Steganalysis of Boundary-based Steganography using Autoregressive Model of Digital Boundaries", IEEE International Conference on Multimedia and Expo (ICME) 2004.

[3] Shaifali Bhatnagar, Shishir Kumar, Ashish Gupta, "An Approach of Efficient and Resistive Digital Watermarking using SVD", 978-1-4799-3080-7/14/c IEEE 2014.

[4] Maria E. Angelopoulou And Peter Y. K. Cheung, "Implementation and Comparison of the 5/3 Lifting 2D Discrete Wavelet Transform Computation Schedules on FPGAs", Journal of VLSI Signal Processing 2007) 2007 Springer Science + Business Media, LLC. Manufactured in The United State. DOI: 10.1007/s11265-007-0139-5.

[5] Jinal Patel, Ketki Pathak," Implementation of the 5/3 Lifting 2D Discrete Wavelet Transform", © 2014 IJEDR | Volume 2, Issue 3 | ISSN: 2321-9939.

[6] Nainesh Agarwal, Nikitas Dimopoulos, "Power Efficient Rapid System Prototyping Usig Codel; The 2D DWT Using Lifting", 0-7803-9195-0/05/© IEEE 2005.

[7] Nainesh Agarwa, Nikitas Dimopoulos, "Rapidly Prototyping DSP Extensions Using CoDeL: The DWT Using Lifting", 0-7803-8886-0/05/ ©2005 IEEE CCECE/CCGEI, Saskatoon, May 2005.

[8] Mamta Juneja and Parvinder S. Sandhu, "An Improved LSB Based Steganography Technique for RGB Color Images", International Journal of Computer and Communication Engineering, Vol. 2, No. 4, July 2013.

[9] Shamim Ahmed Laskar and Kattamanchi Hemachandran, "High Capacity data hiding using LSB Steganography and Encryption", International Journal of Database Management Systems (IJDMS) Vol.4, No.6, December 2012.

[10] Basant K. Mohanty and Promod K.Meher, "Pipelined Architecture for High-Speed Implementation ofMultilevel Lifting 2-D DWT using 9/7 Filters", 1-4244-0969-1/07/©C IEEE 2007.

[11] K.B.Raja, C.R.Chowdary, Venugopal K R, L.M.Patnaik, "A Secure Image Steganography using LSB, DCT and Compression Techniques on Raw Images", 0-7803-9588-3/05/ © IEEE 2005.

[12] S. M. Masud Karim, Md. Saifur Rahman, Md. Ismail Hossain, "A New Approach for LSB Based Image Steganography using Secret Key", 987-161284-908-9/11/ IEEE 2011.

[13] Saket Kumar, Ajay Kumar Yadav, Ashutosh Gupta, Pradeep Kumar, "RGB Image Steganography on Multiple Frame Video using LSB Technique", 978-1-4799-1819-5/15/© IEEE 2015.

[14] G. Raj Kumar, M. Maruthi Prasada Reddy, T. Lalith Kumar , "An Implementation of LSB Steganography Using DWT Technique", International Journal of Engineering Research and General Science Volume2,Issue6,October-November,2014 ISSN 2091-2730.

[15]Manpreet Kaur and Sheenam Malhotra,, " Review Paper on Digital Image Watermarking Technique for Robustness", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4,Issue 5, pp 948-952, May 2014.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 568 - 577, 2016, ISSN:- 2319–8656

Estimation of Walking rate in Complex activity recognition

Hooman Kashanian
Department Of Computer Engineering
Islamic Azad University
Ferdows, Iran

Saeed Sharif
Department Of Computer Science
Islamic Azad University
Ferdows, Iran

Ralf Akildyz
Department Of Electronic And Computer Science
Hacettepe University
Ankara, Turkey

Abstract: Physical activity recognition using embedded sensors has enabled by many context-aware applications in different areas. In sequential acceleration data there is a natural dependence between observations of movement or behavior, a fact that has been largely ignored in most analyses. In this paper, investigate the role that smart devices, including smartphones, can play in identifying activities of daily living. Monitoring and precisely quantifying users' physical activity with inertial measurement unit-based devices, for instance, has also proven to be important in health management of patients affected by chronic diseases, e.g. We show that their combination only improves the overall recognition performance when their individual performances are not very high, so that there is room for performance improvement. We show that the system can be used accurately to monitor both feet movement and use this result in many applications such as any playing. Time and frequency domain features of the signal were used to discriminate between activities, it demonstrates accuracy of 93% when employing a random forest analytical approach.

Keywords: Complex activity recognition; Mobile and ubiquitous environment; Accelerometer; Cell Phones; Humans; Monitoring; Ambulatory, random forest, Online prediction.

1. INTRODUCTION

OSTEOARTHRITIS (OA) is a degenerative disease causing pain, joint stiffness, loss of function and disability [1]. The knee is one of the most commonly affected joints disabling a large proportion of the adult population over a range of daily activities [2].

Exercise is recognized as a key component in the management of knee OA [3] but its effectiveness in restoring joint function is hampered by a lack of individualized programs and by low treatment fidelity. Research studies have quantified the effect of different treatment options in reducing pain and disability reporting small to moderate effects over control groups [4]-[8]. However, the delivered exercise therapies were not tailored to patients' specific impairments or their aspirations, and this may be why none of the studies have reported a definite impact of exercise on quality of life and functional outcomes. Furthermore, the long-term impact of exercise on cessation of the intervention is frequently lost or significantly reduced or simply not reported [8], [9]. This indicates a paucity of longitudinal studies into the effect of optimal rehabilitative approaches and even fewer studies addressing how to optimize the short and long-term exercise compliance in this population group.

Individualized programs can be obtained based on objective measures of patients' joint functional status; however, the routine collection of these measures is rare with the output seldom accessible or made meaningful to healthcare professionals. Simple solutions to enhance compliance may be achieved by solving organizational and accessibility issues (e.g. location, time, work and other commitments) and addressing cost concerns. Furthermore, providing patients with marker of performance and ensuring a correct understanding of the content of rehabilitation will keep them motivated while supporting self-management [10], [11].

It is expected that by prescribing patients exercise regimes based on sound biomechanical assessed deficits and providing them with targets and feedback on performance will enhance compliance and hence treatment effectiveness [12], [13]. Objective measures of knee functional status, referring mainly to knee 3-D angles, are generally obtained in laboratories using expensive, time consuming and difficult to operate equipment. Moreover, the retrievable information is related to an artificial environment over a short period of time. On the other hand, the clinical benefit for long-term monitoring of patients in everyday situations has been advised and it has been proposed that it should be used to inform treatment [14], [15]. Long-term monitoring within each patient environment can only be possible with the use of an ambulatory monitoring system. However, to be effective, this technology needs to be able to inform clinicians on patients' joint status and, be simple and easy to use for patients and allow them to gain feedback on their performance. Despite the use of wearable technology and particularly inertial measurement units (IMUs) gaining popularity within the research environment, clinical uptake remains poor [16], [17].

The main advantage of using wearable devices over standard laboratory-based motion analysis systems to track joint movement relates to the portability of the instrument allowing for prolonged data collection in more realistic environments. However, their everyday use is still limited by poor patient acceptance. To obtain knee angles from IMUs two devices have to be positioned on the shank and thigh of the subject to extrapolate the relative movement between the two segments. The output accuracy is affected by drift from required integration of acceleration and angular velocity values and, artefacts errors due to skin movement and misalignments [18], [19]. In addition to this, their use still requires a certain level of expertise that can limit wide adoption especially in the ageing population. More simple activity monitors based on accelerometry are common and readily accessible on the market for a range of applications. However, the measures obtained are frequently limited to how active a patient is, and few are able to discriminate between activity performed, or able to record step counts and distance travelled. Although important for general activity levels, these parameters do not represent clinically relevant measures directly related to knee joint status. For rehabilitative purposes, it would be important to be able to monitor knee function (e.g., Knee kinematics).

Within our group we explored the use of a flexible conductive polymer material as a sensing modality for knee movement [20]. Laboratory experiments were conducted to evaluate the polymer sensor in measuring flexion and extension angles of the knee in a controlled environment where the knee movement was restricted and standardized with a dynamometer. A subject specific algorithm was defined to obtain measures of knee flexion and extension angles to an accuracy of 1° with the gold standard [20]. The previous study characterized the sensor and validated it in a controlled laboratory setting, but no investigations were conducted to evaluate the sensor's response to free, unconstrained movement. With the intended use of the sensor for knee rehabilitation in the home and clinics, further testing is required to evaluate the sensor capability to follow knee movement patterns in dynamic real life conditions. It was also essential that the sensor had low power requirement to facilitate continuous data acquisition.

The aim of this study was to investigate the reliability of the response of the sensor to everyday tasks and to evaluate its potential towards assessing joint range of motion and activity identification. To support out-of-the laboratory assessments, wearable electronics were developed in the form of a sensing node to allow wireless data acquisition from the sensor. Design constraints included the need for the system to be unobtrusive, low cost, low power and simple to use. This paper focuses on the evaluation and exploitation of the system in reference to the output from a flexible polymer sensor embedded in a pair of leggings. The main contributions relate to the system ability to demonstrate activity discrimination based on a single passive polymer sensor and simultaneously derive a surrogate of knee range of motion from the sensor output to comprehensively describe knee functional status during a specific activity context.

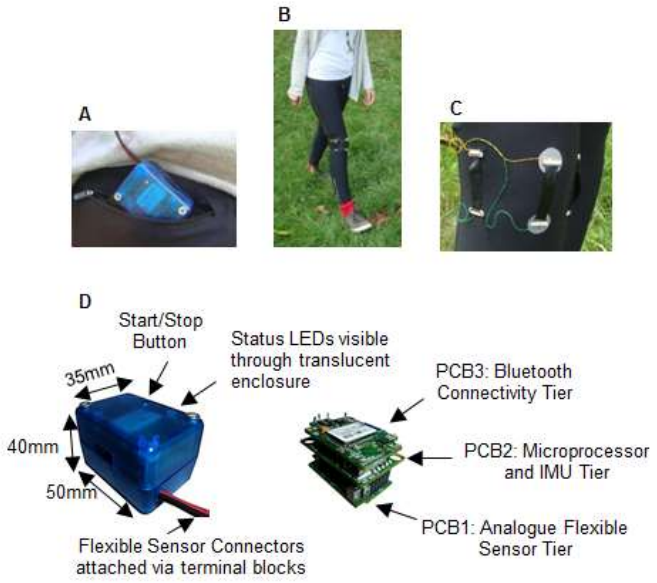


Fig.1. Photographs showing (A) wireless node positioned in the back pocket, (B) sensor integration to a pair of leggings and (C) detail of flexible sensor unit. (D) Photos showing node package with overall dimensions (left) and assembly of printed circuit boards (right).

Two exploitation cases were considered: (i) one that necessitates a subject specific calibration, based on a simplified approach to activity discrimination and (ii) an approach that eliminates the individual calibration but incorporates computational resources for a machine learning approach. Individual subject activity discrimination was successfully achieved based on an innovative combination of two spectral features, median frequency and total power of the spectrum, while group classification was achieved with high accuracy based on a random forest algorithm. Independently of the exploitation set-up of choice, by proving the capability of the system in monitoring knee function in everyday life scenarios, with an appropriate feedback interface, it will be a valuable tool to support knee rehabilitation by providing objective measure of function to clinicians as well as enhancing long term patients' compliance and promoting self-management.

2. METHOD

2.1 Smart Leggings

The sensor unit (Fig.1.C) consists of a conductive flexible polymeric material in the form of a thin (0.2 mm) rectangular strip (50 mm x 100 mm). The conductivity is provided by the presence of graphitized carbon black Nano powder particles (< 500 nm) in a polyurethane substrate. The ratio between the two compounds in the conductive polymer composite is 20:80. Two connectors were attached at each end of the sensor unit. The sensor was secured on to a pair of commercially available leggings (92% polyester, 8% elastin) (Fig.1.B), in a pre-stretched condition, to coincide with the anterior aspect of the knee joint. The composite material has a resistor like-function so when stretched, it changes resistance. Knee motion stretches the sensor allowing for a direct sensing modality for knee flexion/extension movement.

2.2 Data Acquisition: Multi Sensors Wireless Platform

Data from the sensor unit were acquired by means of a custom wireless sensing node (Fig.1.A, D). The developed sensor node consists of three printed circuit board (PCB) tiers (Fig.1.D), each with its own functionality as follows:

i. PCB 1: analogue interface tier accommodating circuitry for the flexible sensor unit; a Wheatstone bridge configuration is used to detect resistive changes within the flexible sensor, the signal is then further amplified by a micropower precision instrumentation amplifier (LT1789, Linear Technology, Milpitas, CA, USA) before being converted to digital values;

ii. PCB 2: core tier with a microprocessor (64MHz PIC18F family, Microchip Technology Inc., Chandler, AZ, USA) and an inertial measurement unit (IMU) embedding a 3 axis accelerometer (ADXL345, Analog Devices Inc, Norwood, MA, USA) and 3 axis gyroscope (L3G4200D, STMicroelectronics, Geneva, Switzerland);

iii. PCB 3: connectivity tier incorporating a small form factor, low power Bluetooth module (RN42, Microchip Technology Inc., Chandler, AZ, USA) allowing wireless data transmission for distances up to 20 meters. Data were acquired synchronously from the IMU and flexible sensor unit at 122 Hz sampling frequency;

The PCB tiers are encased in a box with sides of 35 mm x 50 mm x 40 mm (width x length x height). The node operates on a 3 V battery and its overall mass is 54g. During testing the node was placed on the back pocket of the leggings. Thin wires sewn along the seam of the leggings connected the sensor unit to the PCB 1 of the wireless node.

2.3 Participants

Twelve healthy subjects with no reported knee pain (Age: 27 ± 5 years, Height: 1.7 ±0.1 m, Body Mass: 66 ± 12 kg) took part in the study. The sample size was defined in accordance with earlier recomendations [21]. For a power of 80% and to achieve a specificity of 95%, and assuming test-retest reliability of 0.9 for the sensor outputs with two observations, a sample size of 12 would suffice to allow for observations of test-retest reliability of 0.6 or greater. Written informed consent was obtained from all subjects prior testing, following attainment of ethical approval. Ethical approval was granted by the Imperial College Ethics Research Committee.

2.4 Experimental Procedures

Each participant was tested on two separate occasions with at least a one-week gap between sessions, referred in the text as Test 1 and Test 2. A test session consisted of the participant walking and running both indoors and outdoors, and going up and down consecutive flights of stairs. The indoor test took place along a 30 m corridor; and participants were asked to walk and run this distance 10 times. The stair test was conducted in a public building back stair case using 5 consecutive flights of stair with 10 steps each (width 30cm, height 16cm) with subjects being requested to go up and down the stairs two times. This allowed 10 data sets for both ascending and descending the stairs. The outdoors test was conducted in a quiet nearby park and subjects were instructed to walk and run without stopping for two minutes, twice, with sufficient rest periods allowed between tests. Each subject performed the different activities at their preferred, comfortable speed. During each session, participants were asked to wear the smart leggings and to position the sensor unit to

cover the anterior aspect of their right knee. This imitates the use of the system in home environments where users are unsupervised allowing to evaluate the system in real condition. The sensing node was positioned by the investigator in the back pocket of the leggings once Bluetooth connection was established with a notebook (HP Mini 5103 Notebook PC, Hewlett-Packard Company, Palo Alto, CA, USA) for data acquisition. A test session lasted approximately 45 minutes.

2.5 Data Pre-Processing and Sensor Output

The wearable system allows for simultaneous multisensor data collection, but for the aim of the current study only the flexible sensor unit output was analyses. The use of accelerometer and gyroscopes data is already well established and widely accepted for activity monitoring whereas, the novelty of the present study resides in the ability to provide direct information of knee function while characterizing activities performed using a single passive polymer sensor. Data were pre-processed by filtering and having the DC offset removed from the signal output. A 4th order Butterworth filter with 10Hz cut-off frequency was used. Time histories of the signal outputs were analysed to investigate the capability of the sensor to monitor dynamic knee movement. The range of the signal output in the time domain was evaluated. This range can be considered as a surrogate of the knee range of motion, since the sensor stretches as the knee bends, generating the output which allows mapping the knee flexion/extension movement. Range was normalized to each subject's leg length [22]. Test-retest reliability of the signal output range was assessed by mean of intra-class correlation as defined by Shrout and Fleiss [23] and examined accordingly to the classification of Landis and Kock [24]. Bland and Altman tests statistics [25] were performed to provide a measure of agreement between tests. All data processing and statistical analysis were completed with Matlab (The MathWorks Inc., Natick, MA, USA) and SPSS (SPSS Inc., Chicago, IL, USA) software.

2.6 Spectral Domain Activity Discrimination

A frequency domain approach was adopted to discriminate between activities. A single-sided power spectral density (PSD) analysis was performed using the periodogram method over the whole signal recorded per trial. From the PSD function ($S(f)$), total power of the spectrum and median frequency (MDF) were computed. The total power of the spectrum (P) is the cumulative power of the signal:

$$P = \int_0^{\infty} S(f)df \quad (1)$$

The median frequency is the frequency dividing the signal power spectrum into two equal halves:

$$\int_0^{MDF} S(f)df = \frac{1}{2}\int_0^{\infty} S(f)df \quad (2)$$

These two parameters were used as discriminative features to classify tasks performed. This approach was taken to verify if a simple discriminative algorithm using only these two parameters would allow activity differentiation instead of using computationally complex algorithms involving machine learning techniques.

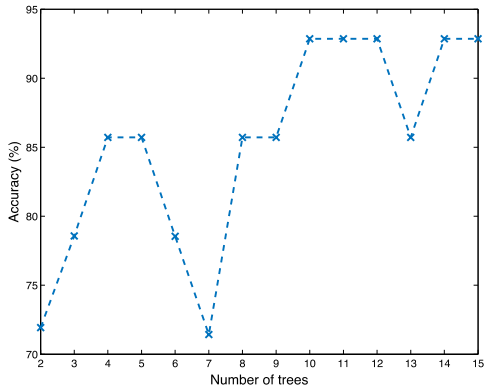


Fig.2. Graph showing the relationship between accuracy and number of trees. descent (upper right), running (lower left) and, walking (lower right).

Table 1

TEST-RETEST RELIABILITY AND BLAND AND ALTMAN TEST RESULTS

	Test-Retest Reliability		Bland and Altman Test				
	ICC Coefficient	95% CI (Lower,Upper Bound)	$\bar{d}$ (mV/m)	$SD_{\bar{d}}$ (mV/m)	$SE_{\bar{d}}$ (mV/m)	Repeatability Coefficient (mV/m)	95% LOA (Lower,Upper Bound)
Run Indoors	0.958	0.860; 0.988	11.0	11.2	3.2	21.9	-10.9; 32.9
Run Outdoors	0.984	0.945; 0.995	1.0	11.8	3.4	23.1	-22.1; 24.1
Walk Indoors	0.897	0.657; 0.970	-0.5	8.9	2.6	17.5	-18.0; 17.0
Walk Outdoors	0.958	0.861; 0.988	-0.9	4.2	1.2	8.3	-9.2; 7.4
Stair Ascent	0.867	0.557; 0.961	-3.6	14.4	4.2	28.2	-31.9; 24.6
Stair Descent	0.938	0.796; 0.982	-0.6	9.3	2.7	18.1	-18.8; 17.5

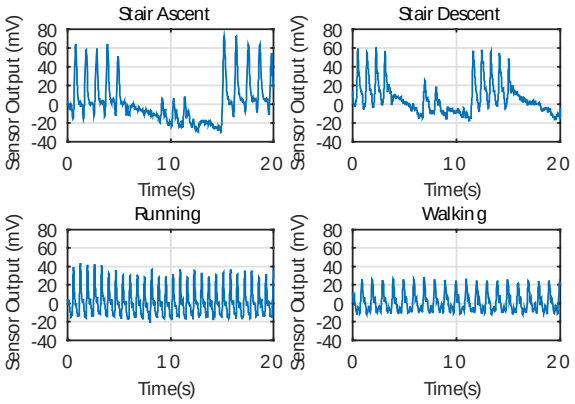


Fig.3. Sensor output time histories during stair ascent (upper left), stair

The sensor capability in discriminating activity was analysed treating participants individually as well as a group for the two tests conducted. This was done to investigate if a general activity detection algorithm could have been defined based on a simple classification method that would not require subject specific calibration (e.g.: identify specific thresholds for each subject to allow real time classification). Median frequency and total power of the spectrum were normalized to each individual's anthropometric features (leg length, body mass and height) when comparing data across participants using the method proposed by Hof [22]. Data analysis was performed using Matlab software. The data showed the need for subject specific calibration when only MDF and power of the spectrum were used for activity classification as no general-purpose thresholds could be defined that would have satisfied all participants' data. Machine learning was then utilized to tackle this problem and overcome the necessity of a baseline subject calibration.

2.7 Random Forest Activity Classification

A Random Forest [26] was used to develop a generalized

classification method to discriminate between activities based on features extracted from the flexible sensor output. Random forests are statistical modern machine learning techniques that allow accurate classification of large datasets that are screened by independent trees, in this instance, classification trees, which form the forest. Each tree develops upon a set of rules based on discriminatory features randomly selected from measured parameters. Random forests perform feature selection automatically to develop each tree that can alternatively be expressed as set of rules. In each node of a tree, a decision is made based on one feature. The random forest combines the response of each tree via majority voting to obtain the ultimate classification response.

The random forest employed in this study is an ensemble of 10 classification decision trees. The number 10 was decided by verifying that increasing the number of trees did not affect

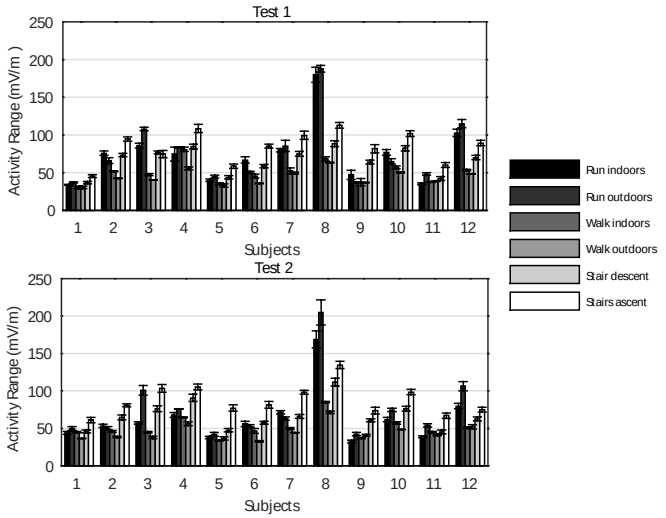


Fig.4. Sensor output range as surrogate of knee range of motion during the different activities for both tests conducted.

accuracy significantly (Fig.2) but, increased the computational complexity of the method. Our aim is to keep the complexity of the data processing to a minimum to allow a timely real time data visualisation in the future.

The ensemble classifies activities into walking, running, and ascending and descending stairs. The ensemble was provided with features from time and frequency domain analysis of the sensor output namely: MDF, power of the spectrum, peak frequency, maximum spectral amplitude and output range of the signal in the time domain. Anthropometric parameters, gender, age, height and leg length, were also utilised. 90% of the data were randomly selected and used for the construction of the trees and 10% of the data were used to test the algorithm.

Performance metrics consisting of accuracy, specificity, sensitivity, and F measure were computed from the confusion matrix to evaluate the classification method. This analysis was performed using Matlab Statistics Toolbox.

3. RESULTS

Typical time series of the sensor output are plotted in Fig.3 for the different activities performed, 20 s time frames are depicted. These plots show that the sensor is able to follow the knee movement during dynamic tasks by capturing the knee flexion/extension repetitions throughout the trials. The output, presented in mV, can thereby be considered a surrogate of knee sagittal kinematics.

The range of the measured voltage from the sensor is shown in the bar charts in Fig. 4 for both tests conducted for each subject. No statistical significant differences were found within subject (p > 0.05). This range could be considered a surrogate of knee range of motion as it quantifies the amount the sensor has stretched due to knee movement during each performed task.

An almost perfect test-retest reliability (ICC > 0.8) was obtained for the output range among all participants (Table I). Bland and Altman test results in Table I and Fig.5 demonstrate good to high agreement between tests with the majority of

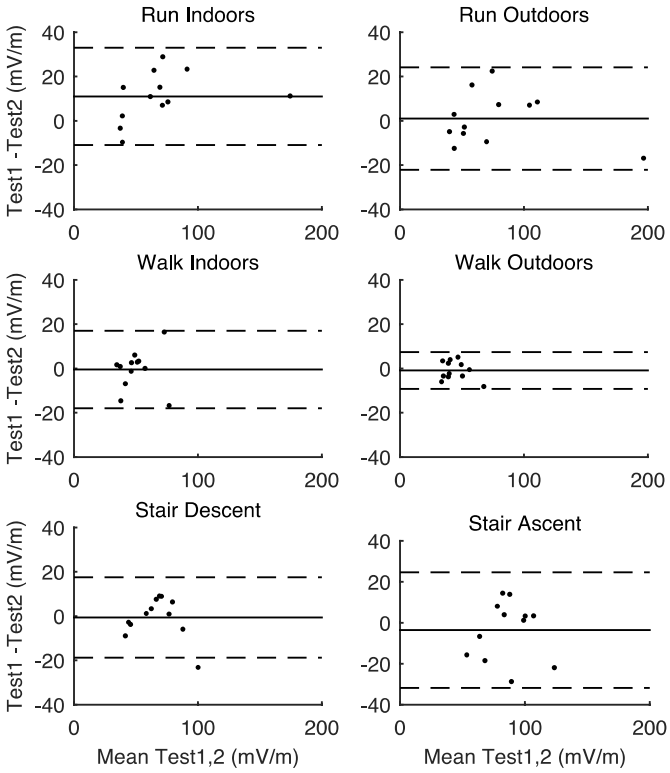


Fig.5. Bland and Altman plot of agreement between Test 1 and 2. Dashed lines represent upper and lower limit of agreement and the solid line represents the mean difference.

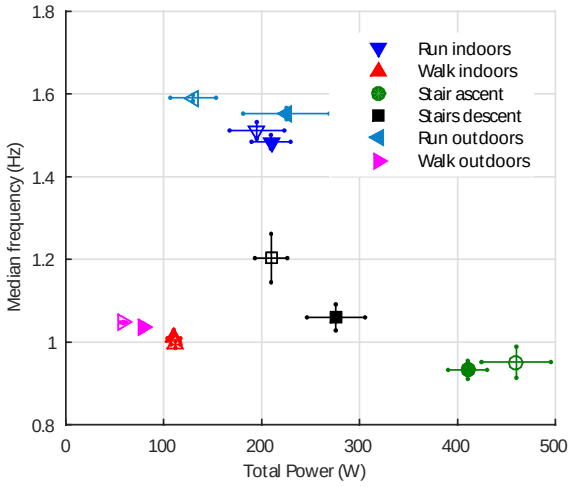


Fig.6. Activities discrimination using MDF and Total Power of the spectrum. Filled markers refers to Test 1, unfilled ones to Test 2 for one participant. The bars indicate ± one standard deviation.

data points falling within the locus of agreement. The highest biased (d^{-}=11 mV/m) was observed for running indoors indicating higher variations occurred in this task. This could be related to the intrinsic variability of the movement but also to the fact that the sensor may have been prone to major movement artefacts with respect to the underlying knee during this fast task which was repeated 10 times.

Fig.6 shows an example of activity clustering for one participant when only using MDF and the total power of the spectrum as discriminative features. Comparable results are obtained for Test 1 and 2. Similar clustering was observed for other participants; summary values of normalized MDF and total power of the spectrum (with standard deviation indicated in brackets below each value) are shown in Table 2.

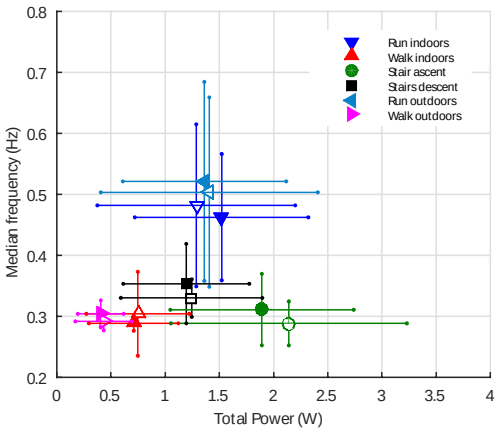


Fig.7. Activities discrimination using MDF and Total Power of the spectrum. Filled markers refers to Test 1, unfilled ones to Test 2 for all participants. The bars indicate ± one standard deviation. Data are normalised to subject specific anthropometric parameters.

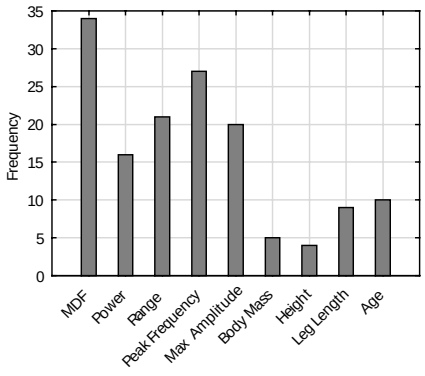


Fig.8. Histogram showing the frequency of the features selected by the Random forest for its decision trees.

Considering the participants as a group led to the discriminatory ability of the MDF and power of the spectrum to be lost (Fig. 7). This

occurred despite normalization of the outputs to subject specific anthropometric parameters [22].

Subject specific calibration is therefore required if these two parameters are to be used alone to identify activities and define thresholds boundaries for the different activities performed.

The possibility of using a machine learning approach, in particular a random forest algorithm, was investigated to allow a generalized discrimination between activities from the sensor output, avoiding the need for subject specific calibration. An ensemble of 10 trees was created using 9 features (Table 2).

Table 2

RANDOM FOREST PERFOMANCE METRICS AND FEATURES

Accuracy	92.8%			
	Run	Walk	Stair Descent	Stair Ascent
Sensitivity	1	1	0.75	NaN
Specificity	1	1	1	0.92
F-score	1	1	0.85	0
# of features	9			
Features	Median Frequency, Total Power of the Spectrum, Peak Frequency, Max Amplitude, Output Range, Body Mass, Height, Leg Length, Age			

Moreover, data from the two tests were combined to allow for more data during the training phase of the random forest. The ensemble, when tested with the remaining 10% of the sample data, performed well with an accuracy of 93%. This implies 93% of the time activities were correctly identified. Detailed performance metrics of the ensemble are shown in Table III.

Fig.8 shows the frequency of occurrence of features automatically selected by the decision trees (i.e.: how many times each feature is picked across all nodes for all trees). Among these features MDF is the most important one, whereas anthropometric parameters, and particularly body mass and height, do not play an important role in the classification process.

4. DISCUSSION

A novel wearable system has been presented that allows simultaneous estimation of a surrogate for knee range of motion and identification of activity type performed. The sensor unit was able to reliably detect knee movement during dynamic activities at different speeds as shown in Table I. The excellent reliability demonstrated the sensor is not affected by movement artefacts allowing for valid results despite don-doff of the system and participants positioning it. This offers the potential of the system to support rehabilitation however, without further work to understand the relationship of the surrogate to knee range of motion it may have limited utility as an outcome measure at this stage.

The time series waveforms of the sensor output (Fig.3) recall typical knee kinematics curves reported in the literature [27], [28]. This indicates the potential to use the sensor output as a surrogate for knee sagittal kinematics, as the output is the direct response to stretching caused by knee flexion/extension movements, thereby permitting acquisition of data from unconstrained environments over extended periods of time. It follows that the range of motion required to perform activities of daily living can also be inferred from the sensor output. Repetitive patterns (Fig.4) were observed among participants in the sensor output range reflecting the knee joint angulation expected for the activities performed [27], [28], lower values indicate the sensor has been stretched less responding to the demand of the activity in requiring less knee flexion/extension. The findings show that walking was the activity that required the smallest range of movement (smaller stretching of the sensor) whereas stair ascent the one with the greatest range of knee motion (greater stretching of the sensor) in agreement with range patterns reported in biomechanical studies [27], [28]. For three of the participants tested, however, running showed the highest stretching span as can be observed from Fig.4, this may be due to the fact that these participants were recreational sport runners and this may be associated to a greater knee flexion/extension range of motion [29]. The knee range of motion is expressed in mV for this preliminary investigation as a first step to identify the capability of the sensor to track knee movement dynamically; the next step will be to identify the relation between the sensor output (mV) and knee angles (°) captured through a 3-D motion analysis system to allow the representation of the output in degrees. However, the possibility to use the output in mV as representation of knee sagittal angles will be explored further together with the clinical interpretation. A database of healthy knees movement, monitored in mV, can be acquired to allow for comparison with pathological knees in the future or, similarly, if we have a baseline measure of a patient knee angles in mV based on the sensor data and having proved, in this study, the accuracy and repeatability of the sensor outputs, the sensor can be used to monitor knee function over time as relative comparison to each individual baseline measure. This also aligns with the idea that functional improvements are relative and specific to each subject. Therefore, there exists a situated use for the sensor in monitoring knee movement also if expressed in mV.

Knee range of motion is often evaluated during the clinical assessment of patients with knee OA with the use of goniometers via a static end range of motion passive test and has frequently been reported as clinically significant parameter in studies of the knee OA population [30]. However, such data and the majority of research data are one off measurements performed within a laboratory or clinical environment and as such not representative of everyday tasks in real life settings. In this study, differences in participants' performances could be appreciated between a task performed indoor or outdoor. An improved understanding of knee function, in real life contexts would permit more effective evaluation of a patient's functional limitations

that could be used to prescribe targeted exercise regime to improve specific functions and follow-up patients' progresses. This may be facilitated by the described system. The system allows continuous long-term monitoring of the knee, which can be expressed as surrogate of knee range of motion, and furthermore it allows the context of the activity to be identified accurately.

Firstly, a simple classification method using MDF and total power of the spectrum was investigated for the identification of activity. This proposed method prioritized the simplified approach (based on only two features) despite the need for subject specific calibration. Different aspects of a signal, and generally of an acceleration signal both in time and frequency domains have been explored to detect activities [31], [32], some of which requiring computationally intense algorithms. MDF discriminatory ability was proposed before for the analysis of acceleration data [33]. MDF alone would not suffice for discrimination between activities using the proposed sensor output; thereby, in this study, it was used in conjunction to the total power of the spectrum. These two features were chosen as they incorporate significant discrimination capability. Good activity discrimination was achieved: data formed defined classes accordingly to the activity performed (Fig.5). This was particularly evident when the participants were analyzed separately, on a subject-by-subject basis. Net separation between activity classes was not achieved when data for all subjects were treated together, as a group, indicating the need for subject specific calibration (Fig. 7). Although sensors could be calibrated for each subject, this may represent a limitation for future clinical adoption, as an additional step is required before actual use, implying extra economic and time costs. This was resolved by successfully employing a random forest algorithm to automatically detect activities in a generalised fashion. This method was mathematically more complex but has the advantage that can be applied without the need for subject specific calibrations.

Machine learning techniques, among which random forests, have been recently used to classify activities from acceleration data acquired via a number of 3-axis accelerometers or smartphone positioned on different parts of the body [31],[32], [34]-[39]. Most of the studies involved the simultaneous use of two or more devices in different positions to increase the accuracy of the classification methods proposed. This leads to a bulkiness of the system not compatible with patients' preferences [40]. On the other hand, the use of one sensor alone implied specific positioning on areas that could interfere with activities of daily living (e.g.: chest, bulky phone in the pocket) or more visible to the other (e.g: ear) against patients' discreetness. The classification method proposed showed high accuracy (93%) utilising 9 features from a single sensor alone while allowing discrete data monitoring. The accuracy achieved compares well with the accuracy reported in previously conducted studies (range 80-99%) using more conventional acceleration signals to detect activity. Further improvements in the accuracy may be achieved via investigating a larger set of subjects that covers the pathological case as well.

The feature that played the greatest role in the activity classification was the MDF. All the features utilised allow for an easy implementation. The random forest demonstrated good discrimination ability in correctly identifying activities performed as seen in the performance metrics table (Table III). Among the testing set values none of the data referred to stair ascent thereby explaining the low F score and sensitivity values. A larger data set will be collected to further test the method proposed having demonstrated the viability of the system for activity classification through this study. Also, once the random forest is trained, the identification of activities for future subjects can be achieved in real time. The features utilized will be calculated to allow real-time feedback in an automated fashion by using a moving window method as the data are collected and, not over whole trial as conducted for this study. Visual feedback of the data for patients and clinicians will complement the wireless system to allow an easy and fast interpretation of the data for clinical use. Data will be made available via smartphone/tablet application or in the form of a one-page report on patient progress. Although accelerometers are established systems for activity recognition or activity level quantification in their simplest form, the sensor proposed allows also for range of movement estimation not achievable with one accelerometer. This dual functionality represents an advantage of our system over existing technology. Although the smart leggings utilised for this study still shows visible electronics, these will be integrated into clothing in the next prototype to comply with patients' needs and maximise acceptance [40]. Moreover, the system proposed requires minimum training for the end user to permit independent utilization.

5. CONCLUSION

Findings from this study demonstrate the feasibility of the novel sensing system in monitoring knee movement and classifying activities of daily living. Being able to monitor knee functional status outside laboratory environments will bring great advantage to the rehabilitation of patients with knee OA. Objective measures of knee health can both inform treatment and motivate patients to comply with prescribed rehabilitation regimes to enhance clinical benefit.

Additional activities will be included in further testing to have a more comprehensive classification of activity of daily living and to explore the possibility to express the output in degrees. System design together with a visual feedback tool will be improved to reflect end users preferences, both patients and health professionals, and ultimately progress into clinical adoption. The use of the sensor can also be expanded to the monitoring of clinically used performance tests to assess patients' physical function. A study conducted within our group showed the ability of the sensor to monitor performance during exercises extrapolated from a knee OA rehabilitation class [41]. Assessment of performance-based tests as suggested by OA guidelines could be included as additional processed outcome of the sensor increasing the clinical usefulness of the information obtained from the novel system.

6. REFERENCES

[1] J. J. Guiry, P. van de Ven, and J. Nelson, "Multi-Sensor Fusion for Enhanced Contextual Awareness of Everyday Activities with Ubiquitous Devices," *Sensors*, vol. 14, no. 3, p. 5687, 2014.

[2] C. Giannella, J. Han, J. Pei, X. Yan, and P. S. Yu, Mining Frequent Patterns in Data Streams at Multiple Time Granularities. MIT Press, 2003, ch. 3.

[3] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic Textures," *Int. J. Comput. Vis.*, vol. 51, no. 2, pp. 91–109, 2003.

[4] Y. Hanai, J. Nishimura, and T. Kuroda, "Haar-like filtering for human activity recognition using 3d accelerometer," in *Proc. IEEE 13th Digit. Signal Process. Workshop, 5th IEEE Signal Process. Educ. Workshop*, Jan. 2009, pp. 675–678

[5] Guiry, J.J.; van de Ven, P.; Nelson, J. Multi-Sensor Fusion for Enhanced Contextual Awareness of Everyday Activities with Ubiquitous Devices. Sensors 2014, 14, 5687-5701.

[6] Tzu-Yi Hung, Jiwen Lu and Yap-Peng Tan, "Graph-based sparse coding and embedding for activity-based human identification," *2013 IEEE International Conference on Multimedia and Expo (ICME)*, San Jose, CA, 2013, pp. 1-6.

[7] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher Discrimination Dictionary Learning for Sparse Representation," in *Proceedings of the 2011 International Conference on Computer Vision*, 2011, pp. 543–550.

[8] L. Liu, C. Shen, L. Wang, A. van den Hengel, and C. Wang, "Encoding High Dimensional Local Features by Sparse Coding Based Fisher Vectors," Nov. 2014.

[9] R. Raina, A. Battle, H. Lee, B. Packer, A. Y. Ng, Self-taught learning: transfer learning from unlabeled data, in: Proc. Int. Conf. on Machine Learning (ICML), 2007.

[10] S. Bhattacharya, P. Nurmi, N. Hammerla, and T. Plötz, "Using unlabeled data in a sparse-coding framework for human activity recognition," *Pervasive Mob. Comput.*, vol. 15, pp. 242–262, 2014.

[11] V. Kumar, C. Narasimham, and B. Sujith, "Classification of Time Series Data by One Class Classifier using DTW-D," *Procedia Compute. Sci.*, vol. 54, pp. 343–352, 2015.

[12] R. J. Martin, "A metric for ARMA processes," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1164–1170, Apr. 2000.

[13] Z. Dong, W. Liang, Y. Wu, M. Pei, and Y. Jia, "Nonnegative correlation coding for image classification," *Sci. China Inf. Sci.*, vol. 59, no. 1, pp. 1–14, 2016.

[14] A. Ravichandran, R. Chaudhry, and R. Vidal, "Categorizing Dynamic Textures Using a Bag of Dynamical Systems," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 35, no. 2, pp. 342–353, Feb. 2013.

[15] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition.," *Sensors (Basel).*, vol. 14, no. 6, pp. 10146–76, Jan. 2014.

[16] M. Shoaib, S. Bosch, H. Scholten, P. J. M. Havinga and O. D. Incel, "Towards detection of bad habits by fusing smartphone and smartwatch sensors," *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, St. Louis, MO, 2015, pp. 591-596

[17] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors.," *Sensors (Basel).*, vol. 16, no. 4, p. 426, Jan. 2016.

[18] S. G. Trost, Y. Zheng, and W.-K. Wong, "Machine learning for activity recognition: hip versus wrist data," *Physiol. Meas.*, vol. 35, no. 11, p. 2183, 2014.

[19] M. Stikic, D. Larlus, and B. Schiele, "Multi-graph Based Semi-supervised Learning for Activity Recognition," in *2009 International Symposium on Wearable Computers*, 2009, pp. 85–92.

[20] M. Stikic and B. Schiele, "Location and Context Awareness: 4th International Symposium, LoCA 2009 Tokyo, Japan, May 7-8, 2009 Proceedings," T. Choudhury, A. Quigley, T. Strang, and K. Suginuma, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 156–173.

[21] H. Liu, L. Yu, W. Wang, and F. Sun, "Extreme learning machine for time sequence classification," Neuro computing, vol. 174, Part A, pp. 322–330, 2016.

[22] J. R Munkers, Topology vol. 2: Prentice Hall Upper Saddle River, 2000.

[23] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[24] H. Ney, D. Mergel, A. Noll and A. Paeseler, "A data-driven organization of the dynamic programming beam search for continuous speech recognition," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.* 1987, pp. 833-836.

[25] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher Discrimination Dictionary Learning for Sparse Representation," in *Proceedings of the 2011 International Conference on Computer Vision*, 2011, pp. 543–550.

[26] L. Liu, C. Shen, L. Wang, A. van den Hengel, and C. Wang, "Encoding High Dimensional Local Features by Sparse Coding Based Fisher Vectors," Nov. 2014.

[27] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse Representation Based Fisher Discrimination Dictionary Learning for Image Classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, 2014.

[28] G. Ciuti, L. Ricotti, A. Menciassi, and P. Dario, "MEMS Sensor Technologies for Human Centred Applications in Healthcare, Physical Activities, Safety and Environmental Sensing: A Review on Research Activities in Italy," Sensors, vol. 15, no. 3, p. 6441, 2015.

[29] Debraj De, Pratool Bharti, Sajal K. Das, Sriram Chellappan, "Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare", IEEE Internet Computing, vol.19, no. 5, pp. 26-35, Sept.-Oct. 2015, doi:10.1109/MIC.2015.72

[30] C. Torres-Huitzil and A. Alvarez-Landero, "Mobile Health: A Technology Road Map," S. Adibi, Ed. Cham: Springer International Publishing, 2015, pp. 147–169.

[31] U. Fareed, "Smartphone Sensor Fusion Based Activity Recognition System for Elderly Healthcare," in Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare, 2015, pp. 29–34.

[32] Ryan M. Gibson, Abbes Amira, Naeem Ramzan, Pablo Casaseca-de-la-Higuera, and Zeeshan Pervez, "Multiple Comparator Classifier Framework for Accelerometer-Based Fall Detection and Diagnostic," Applied Soft Computing, 2015.

[33] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey," in Architecture of Computing Systems (ARCS), 2010 23rd International Conference on, 2010, pp. 1–10.

[34] G. Fortino, X. Li, X. Lin, O. Mayora, E. Natalizio, and M. R. Yuce, "Wireless Technology for Pervasive Healthcare," Mob. Networks Appl., vol. 19, no. 3, pp. 273–275, 2014.

[35] N. D. Lane, M. Lin, M. Mohammod, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A. T. Campbell, and T. Choudhury, "BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing," Mob. Networks Appl., vol. 19, no. 3, pp. 345–359, 2014.

[36] R. I. Ramos-Garcia and A. W. Hoover, "A Study of Temporal Action Sequencing During Consumption of a Meal," in *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, 2013, pp. 68:68–68:75.

[37] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "RisQ: Recognizing Smoking Gestures with Inertial Sensors on a Wristband," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, 2014, pp. 149–161.

[38] Guiry, J.J.; van de Ven, P.; Nelson, J. Multi-Sensor Fusion for Enhanced Contextual Awareness of Everyday Activities with Ubiquitous Devices. Sensors 2014, 14, 5687-5701.

[39] I. Sabek, M. Youssef, and A. Vasilakos, "Ace: An accurate and efficient multi-entity device-free wlan localization system," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 261–273, Feb. 2015.

[40] Z. Sheng *et al.*, "A survey on the ietf protocol suite for the Internet of Things: Standards,.challenges, and opportunities," *IEEE Wireless Common.*, vol. 20, no. 6, pp. 91–98, Dec. 2013.

[41] Ö Yürür, C. H. Liu and W. Moreno, "Light-Weight Online Unsupervised Posture Detection by Smartphone Acc elerometer," in *IEEE Internet of Things Journal*, vol. 2, no. 4, pp. 329-339, Aug. 2015.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 578 - 583, 2016, ISSN:- 2319–8656

An Efficient Approach for Enhancing the Security of Amazigh Text using Binary Tree

Fatima Amounas
R.O.I Group, Computer Sciences
Department, Moulay Ismaïl
University, Faculty of Sciences
and Technics, Errachidia,
Morocco.

Abstract: Now a day's Cryptography is one of the broad areas for researchers. Due to its importance, several cryptography techniques are adopted by many authors to secure the data, but still there is a scope to improve the previous approaches. The main of our research is to develop a novel Approach for enhancing the security of Amazigh Text using binary tree. The plaintext considered is the combination of Unicode characters. This paper contributes in the area of elliptic curve cryptography by encrypting data using matrix approach and using the concept of tree traversal method for enhancing the security of the encrypted points. The security goals were enhanced by making it difficult for attacker to predicate a pattern as well as speed of the encryption/decryption scheme. The results show strength of the algorithm.

Keywords: Elliptic Curve Cryptography, Binary Tree, In-order, Pre-order, Post-order, Unicode, Amazigh Alphabet.

1. INTRODUCTION

Cryptography is the science of protecting information from undesirable individuals by converting it into a form non-recognizable by its attackers while stored and transmitted. Security is a big concern and securing crucial data is very essential, so that the data cannot be change or misused for any illegal purposes. For ensuring the security, the plain text is converted to cipher text by the sender. This process is called encryption. Decryption is exactly reverse process of encryption by which intended user can decode the message to its original form.

Elliptic Curve Cryptography (ECC) is one of the most efficient techniques that are used for ensuring the security, because it is difficult for the adversary to solve the elliptic curve discrete logarithm problem to know the secret key that is used in encryption and decryption processes.

Now a day's Different mathematical schemes and algorithms are there to scuttle the content of the message using ECC technique. Many scientists were doing research on the existing methods to make more strong and unbreakable ciphers by enhancing them [1, 2, 3]. In this paper, an enhanced approach to secure Amazigh text is introduced which is based on binary structure, so that it will more secure and protect the confidentiality and integrity of the information being transmitted. From the literature, the tree is considered as a non linear data structure mainly used to represent the hierarchical relationship between data [4]. Recently, it play a vital role in compiler constructions, operating systems and others system software's.

There are two types of trees: General trees and Binary trees. A general tree is a finite non empty set of nodes and can contain any number of nodes. A binary tree is a finite set of elements that is either empty or is partitioned into three disjoint subsets. The first one contains a single element called the root of the tree. The other two subsets themselves are binary trees called the left sub tree and right sub tree. A binary tree is very useful data structure when two way decisions must be made at each point in a process. This structure is used in our proposed encryption algorithm for enhancing the security and the detailed explanation is presented in section 3.

2. BACKGROUND INFORMATION

In this section we provide some basic details required in the proposed method.

2.1 Binary Trees

A binary tree is a hierarchal data structure and it is a common tree that is used for various practical applications and computational processes. Binary trees are a type of data structures that contain nodes with information attached to these nodes.

The information can be processed in any way such that the nodes in the tree can be traversed from top to bottom or from left to right or right to left or bottom to top or any other possible ways. The nodes in the binary tree can be navigated in many different ways. One such possible way is taken and an encryption and decryption algorithm is proposed using the nodes of these binary trees. A binary tree is a tree where every node has at most degree as 2 and levels are labeled along with the name of the nodes such as leaf nodes and child nodes. Elements can be inserted in the nodes of a binary tree and they can be traversed from one node to another node.

Binary search trees are used for searching elements in binary tree through traversing in different possible ways possible. The root node is distinguished from every other node in a binary tree and all the nodes can be reached from the root node by traversing from the root node. Tree is a restricted form of graph and it does not contain cycles and it comes under the category of acyclic graphs in graph theory and applications.

Tree traversal (also known as tree search) is a form of graph traversal and refers to the process of visiting (checking and/or updating) each node in a tree data structure, exactly once [5, 6]. Trees can be traversed in pre-order, in-order, or post-order.

- *Pre Order Traversal*

In pre-order traversal:

1. Display the data part of the root (or current node).
2. Traverse the left sub tree by recursively calling the pre-order function.
3. Traverse the right sub tree by recursively calling the pre-order function.

Pre-Order Algorithm:

```
preorder(node)
{
if (node = null)
    return;
else
    visit (node)
    preorder (node.left)
    preorder (node.right)
}
```

- *InOrder Traversal*

In In-order traversal:

1. Traverse the left sub tree by recursively calling the In-order function
2. Display the data part of the root (or current node).
3. Traverse the right sub tree by recursively calling the In-order function.

In-Order Algorithm:

```
Inorder(node)
{
if (node = null)
    return;
else
    Inorder (node.left)
    visit (node)
    Inorder (node.right)
}
```

- *Post Order Traversal*

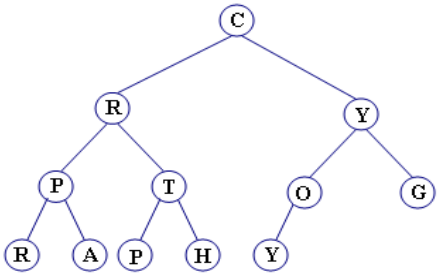
In post order traversal:

1. Traverse the left sub tree by recursively calling the post-order function.
2. Traverse the right sub tree by recursively calling the post-order function.
3. Display the data part of the root (or current node).

Post-Order Algorithm:

```
postorder(node)
{
if (node = null)
    return;
else
    postorder (node.left)
    postorder (node.right)
    visit (node)
}
```

Example: By applying the tree traversal techniques the result is as shown below:



Plaintext: CRYPTOGRAPHY

Inorder: RPARPTHCYOYG

Preorder: CRPRATPHYOYG

Postorder: RAPPHTRYOGYC

2.2 Elliptic Curve

An Elliptic Curve E consists of the set of points (X , Y , Z) that satisfy the following homogeneous Weierstrass equation:

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

Where a_i (i=1, 2, 3, 4, 5, 6) are elements of a finite field [7] and with the exception that the triple (0, 0, 0) is not a point on E.

If we set Z= 0 and substitute $x = X / Z$, $y = Y / Z$ then we get the equation:

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 .$$

The above equation is called the affine Weierstrass equation. If a point P satisfy the homogeneous Weierstrass equation and the equation:

$$\frac{\partial F}{\partial X} = \frac{\partial F}{\partial Y} = \frac{\partial F}{\partial Z} = 0$$

Then we call that point singular and we call the Weierstrass equation also singular, note that singular Weierstrass equations are not of interest in the cryptography [8].

We need now criteria that can help us to determine if a given affine Weierstrass equation singular is or not. The discriminate Δ (field element) is such a tool, which can be defined as follow:

$$d_2 = a_1^2 + 4a_2$$

$$d_4 = 2a_4 + a_1 a_3$$

$$d_6 = a_3^2$$

$$d_8 = a_1^2 a_5 + 4a_2 a_6 - a_1 a_3 a_4 + a_2 a_3^2 - a_4^2$$

$$c_4 = d_2^2 - 24\, d_4$$

$$\Delta = -d_2^2 - 8d_4^3 - 27\, d_6^2 + 9 d_2 d_4 d_6$$

$$j(E) = c_4^3 / \Delta$$

If Δ =0, then affine Weierstrass equation is singular, otherwise not singular [9]. We call j(E) the j-invariant of the elliptic curve E. Note that only elliptic curves E over finite fields are of interest in cryptography [10].

The definition of group of points over elliptic curve E:

1. There is a point $\Omega \in E$, such that for all $P \in E$, $P + \Omega = \Omega + P = P$, (the identity of the group).
2. If $P \neq \Omega$ and $P=(x_1, y_1)$ then -P is $(x_1, -y_1 - a_1 x_1 - a_3)$.
3. If two points on E have same x-coordinate then either P=Q or P=-Q.
4. If Q = -P, then P + Q = Ω.
5. For two points $P \neq \Omega$ and $Q \neq \Omega$ on E, the addition is defined as follows. Draw the line through P and Q to intersect the curve in a third point; then reflect that point in the x-axis.
6. For two points $P \neq \Omega$ and $Q \neq \Omega$ on E, if P = Q, use the tangent line at P. The identity of the group is Ω, the "point at infinity", which conceptually lies at the top and bottom of every vertical line.

The following figure shows the addition of two points over the elliptic curve E:

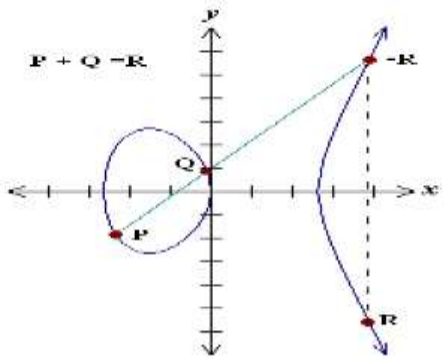


Figure 1. Addition of two points over elliptic curve E.

The discrete logarithm problem over the elliptic curve E is the following: given two points P and Q in a group that satisfy E, find a number α such that $\alpha P = Q$, α is called the discrete logarithm of Q to the base P.

For more information about elliptic curves in cryptography see [11, 12, 13].

2.3 Amazigh Language

In Morocco, Amazigh language is used by tens of millions of people mainly for oral communication, and has been introduced in mass media and in the educational system. Due to its complex morphology as well as to the use of the different dialects: Tarifit in the North, Tamazight in the center and Tashlhit in the southern parts of the country in its standardization, the Amazigh language presents interesting challenges for many researchers [14, 15, 16].

The official graphic system for writing Amazigh is Tifinagh. It does not have capitalization in its script and it is written from left to right. IRCAM uses 33 characters (consisting of: 27 consonants, 2 semi-consonants and 4 vowels). The Figure 2 represents the repertoire of Tifinagh which is recognized and used in Morocco. The total numbers of Tifinagh letters are occupying 2D30-2D7F plage in Unicode. There are 55 defined characters [17].

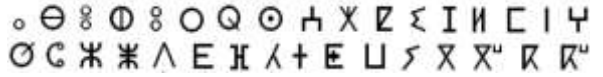


Figure 2. Tifinagh characters adopted by IRCAM.

3. Proposed Approach

The proposed algorithm is an attempt to present a new approach for enhancing the security of Amazigh text based on binary tree in such a way that the new approach can make use of tree traversal method to achieve higher level of security. Let $E_p(a,b)$ be the set of all elliptic curve points over finite field GF(p) corresponding to the defined curve, here $E_p(a,b)$ and the base point P are publicly known [18]. Suppose Alice wants to send a plaintext message to Bob aver an insecure channel, the procedure is as follows:

3.1 Encryption Process

1. Take any sentence Amazigh as input of the algorithm and imbed the given string into respective mapping points on elliptic curve.

2. Convert the given sequence into a data matrix with entries in elliptic curve, called M.

$$\{P_1(x_1,y_1), P_2(x_2,y_2), P_3(x_3,y_3), \ldots, P_n(x_n,y_n)\}$$

$$\Downarrow$$

$$M = \begin{pmatrix} P_1 & P_2 & P_3 & \ldots & P_r \\ P_{r+1} & P_{r+2} & P_{r+3} & \ldots & P_s \\ P_{s+1} & P_{s+2} & P_{s+3} & \ldots & P_t \\ P_{s+1} & P_{s+2} & P_{s+3} & \ldots & P_n \end{pmatrix}$$

Here r=n/4 and s=n/2 and t=3n/4.

If n isn't divided by 4, the points have padded with $\propto$ in order to fill the entries of the matrix M.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 578 - 583, 2016, ISSN:- 2319–8656

3. Construct a key matrix of the same order as the order of the matrix M. The key matrix is denoted K.

4. Multiply the key matrix with the data matrix Q= K×M and insert the resultant values in the binary tree as proposed.

5. Construct a random complete binary tree with total number of nodes n=length (string). Label the nodes starting from the root node in that order as seen in Figure 3.

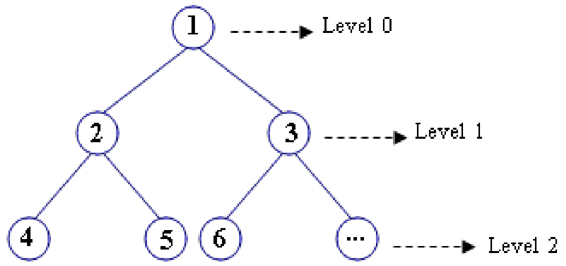


Figure 3. A Complete Binary Tree with two levels

6. Choose a random number integer k and compute a secure key k_1=kP_B.

7. Let b= ($b_j b_{j+1} b_{j+2}$), where j is bit position (LSB→MSB), which decides which traversal method, has to be performed on binary tree.

8. Select the tree traversal method based on the selected digit in the key. Divide the decimal number by 3 and keeps track of the remainder.

If the number is 0 ➔ Pre-order, 1 ➔ In-order and

2 ➔ Post -order.

9. Determine the result of the selected traversal method of the complete binary tree.

10. Insert the resultant values in the binary tree as proposed. Repeat steps 7 to 9 for m times.

11. Send the result cipher text (kP, C_j) to the receiver.

3.2 Decryption Process

Decryption is done by reversing the procedure.

1. Determine the alphabetical representation of the received message and extract kP.
2. Compute k_1= n_B(kP) with n_B is his own private key.
3. Processing the reverse process the various steps and constructs the complete binary tree.
4. Convert each node into point on elliptic curve and insert them in the data matrix, called Q.
5. Compute M = K^{-1}Q to obtain the mapping points.
6. Reverse the embedding to recover the plaintext.

4. RESULTS AND DISCUSSION

4.1 Illustration with an example

Let the message to be encrypted be:

“ⵉⵟⴷⴰ ⵓⵟⵃⵎⴰⴷ ⵉⴷⵃⵉⵟⵏ ⵅⵃ ⵉⵏⵃⵎⴰⴷⵏ ⵅ ⵜⵉⵏⵎⵃ.”

That means:

“The teacher distributed books to students at the school.”

Consider a non-singular elliptic curve defined as follows:

$$y^2=x^3-x+19 \bmod 71.$$

The points on the elliptic curve over E_{71}(-1, 19) are shown below in Figure 4.

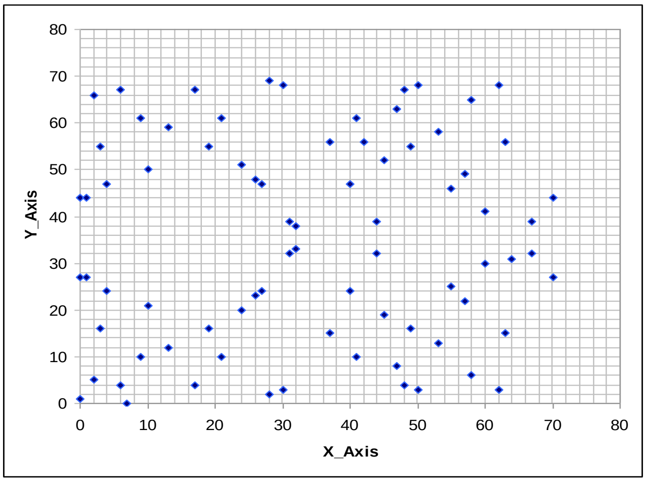


Figure 4. The elliptic curve E_{71} (-1, 19)

The base point P is chosen as (1, 27). Assume that Alice wants to send the above message to Bob. The initial mapped points are given as:

P_i={(17,4) (21,10) (47,8) (63,15) (24,20) (41,10) (3,16) (13,12) (55,25) (63,15) (47,8) (24,20) (17,4) (47,8) (13,12) (17,4) (3,16) (9,10) (24,20) (58,6) (9,61) (24,20) (17,4) (9,10) (13,12) (55,25) (63,15) (47,8) (9,10) (24,20) (26,23) (24,20) (44,32) (17,4) (9,10) (55,25) (13,12)}

These points can be written as a 4×9 matrix denoted M.

The random nonsingular matrix K is chosen as:

$$K=\begin{pmatrix} 2 & 0 & 4 & 3 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 2 & 2 \end{pmatrix}$$

The above set of points is converted into the following Cipher-text through the data matrix approach:

Q_i={(6,67) (10,50) (57,22) (37,15) (60,30) (17,67) (42,15) (67,32) (21,10) (19,55) (41,10) (62,3) (3,55) (13,59) (26,23) (58,6) (70,44) (67,39) (48,67) (44,32) (53,58) (49,16) (17,67) (24,51) (3,16) (1,44) (67,32) (45,52) (49,16) (47,8) (21,10) (53,13) (27,24) (3,55) (42,15) (30,68) (60,30) (37,15) (26,48) (41,10) (17,67) (31,32)}

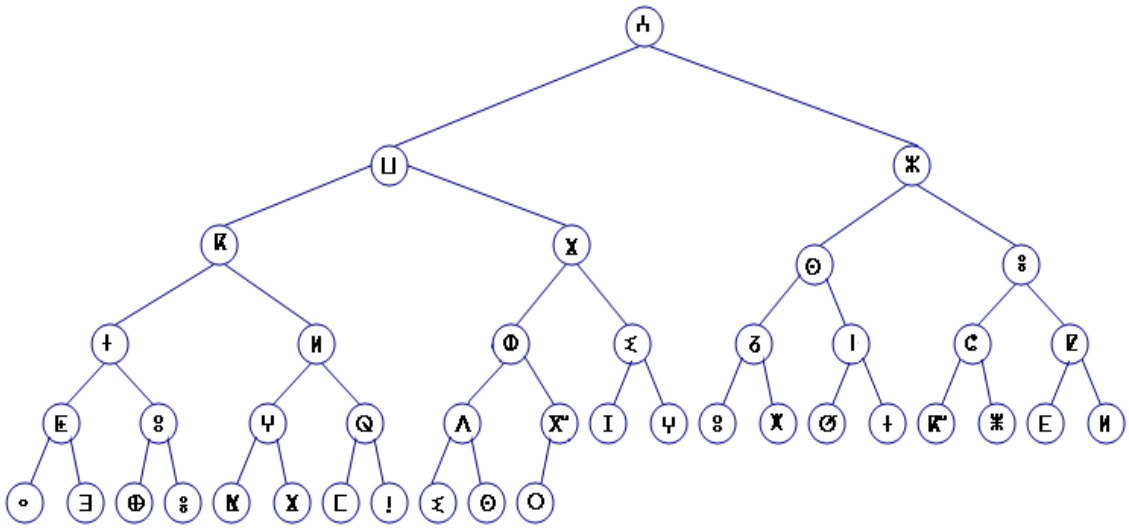
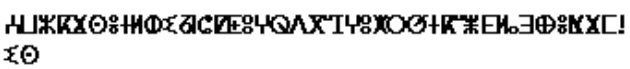


Figure 5. Complete binary tree of the mapping points.

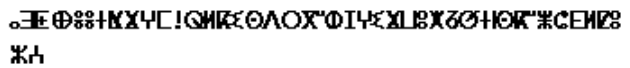
The alphabetical representation of the result sequence is given as:



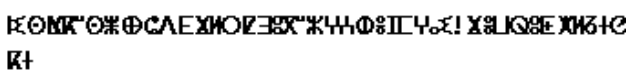
Next, insert the resultant characters in the complete binary tree as seen in Figure 5.

After applying the traversal methods based on the secure key: k_1= (21,10), we get

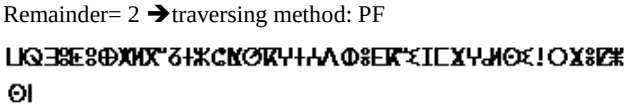
Remainder= 2 ➔traversing method: PF



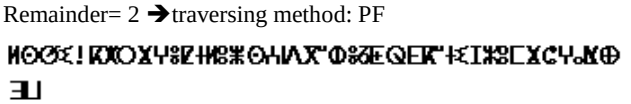
Remainder= 1 ➔traversing method: IF



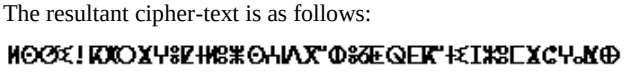
Remainder= 2 ➔traversing method: PF



Remainder= 2 ➔traversing method: PF



The resultant cipher-text is as follows:



At the receiving side, decryption is done by reversing the procedure.

5. RESULTS & DISCUSSIONS

Data Security is a very important aspect. Security of an algorithm is measured by computing number of decryption steps. Higher the number of decryption steps to decrypt the cipher text to get original message shows higher level of security. It is shown that the security can be enhanced by applying the proposed method. For different data sets, results of existing algorithms [19] and [20] are compared with the proposed algorithm. To enhance security, tree traversal method is performed on encrypted data.

The table 1 illustrates the number of decryption steps for input text data of different lengths. As shown below, number of decryption steps varies according to different data values.

Table 1. Number of decryption steps of different algorithms.

Input data size	10	30	50	70
Alg. [19]	27	78	122	195
Alg. [20]	44	95	143	214
Proposed Alg.	63	113	168	231

Graphical representation of the above described table is shown in Figure 6 for computing security in terms of number of decryption steps.

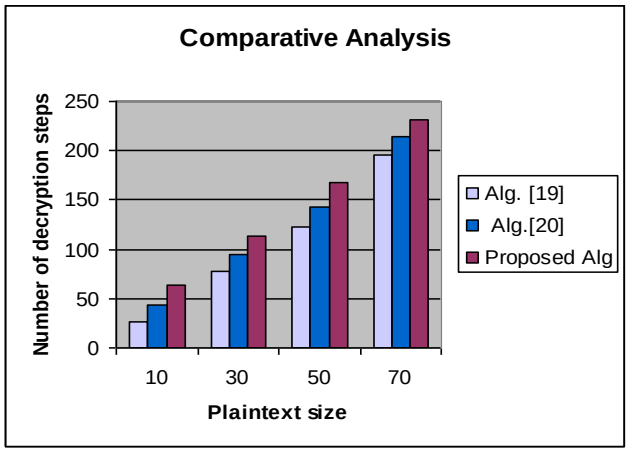


Figure 6. Comparison performances with the existing algorithms.

According to the graph, there is tendency that number of decryption steps of the proposed algorithm, and compared algorithms increases with text data size. According to the proposed algorithm, number of decryption steps taken by decryption algorithm to decrypt the cipher text is high than number of decryption steps of existing algorithms.

6. CONCLUSION

Information security is one of the most important issues in the recent times. ECC is one of the most efficient public key cryptosystems that is secured against adversaries because it is difficult for them to solve the elliptic curve discrete logarithm problem to find the secret key. In this paper, a new efficient approach has been proposed to improve the ECC cryptosystem based matrix. The main contribution is to enhance the security of the proposed method using binary tree. To enhance security, tree traversal method is performed on encrypted data. In this paper the possibility of arranging text into binary tree, and the chose of traversing method provide better performance in this regard. As results, this proposed algorithm can be applied to various Character encoding systems. In near future, it can be applied to various software packages like banking, Educational system etc.

7. REFERENCES

[1] Tanusree Saha, 2015. "An Enhanced Approach to Secure Message Using Combination of Symmetric and Asymmetric Cryptography and Triangulation Method", International Journal of Latest Trends in Engineering and Technology, Vol. 5 Issue 1.

[2] S. Thiraviya Regina Rajam and S. Britto Ramesh Kumar, 2015. "Enhanced Elliptic Curve Cryptography", Indian Journal of Science and Technology, Vol 8 (26).

[3] G.Prabu kanna and V.Vasudevan, "Enhancing the Security of User Data Using the Keyword Encryption and Hybrid Cryptographic Algorithm in Cloud", International Conference on Electrical, Electronics, and Optimization Techniques, IEEE, 2016.

[4] Yedidyah Langsam, Moshe J. Augenstein, M.Tenebaum, 2000. "Data Structures using C and C++", 2nd Edition, 249-319, ISBN-81-203-1177-9.

[5] T. H. Corment, C. Leiserson, R. Rivest and C. Stein, "Introduction to Algorithms", 2nd Edition, MIT Press, September 2001.

[6] Sumit Sharma and Shobha bhatt, 2015. "Encryption of Message Block using Binary Tree in Block Cipher System: An Approach", International Journal of Science Technology & Engineering, Vol. 2, Issue 01.

[7] Nils Gura , Sheueling Chang Shantz , Hans Eberle , Sumit Gupta , Vipul Gupta , Daniel Finchelstein Edouard Goupy, 2002. "An End-to End Systems Approach to Elliptic Curve Cryptography", In Cryptographic Hardware and Embedded Systems, pp. 349-365.

[8] Darrel Hankerson, Alfred Menezes and Scott Vanstone, Guide to elliptic curve cryptography, Springer-Verlag, 2004.

[9] N. Koblitz, 1987. "Elliptic curve cryptosystems". Mathematics of Computation, 48: 203-209.

[10] C. Doche G. Frey T. lange K. Nguyen R. Avanzi, H. Cohen and F. Vercateren. "Handbook of elliptic and hyperelliptic curve cryptography". Chapman and Hall, 2006.

[11] William Stalling, "Cryptography and network security"4th edition, Prentice Hall, 2006.

[12] Vishwa gupta, 2012. "Advance cryptography algorithm for improving data security" Int. J of Advanced Research in Computer Science and Software Engineering, Vol. 2, Issue 1.

[13] Sonali Nimbhorkar1 and Dr. L.G.Malik, 2013. "Prospective Utilization of Elliptic Curve Cryptography for Security Enhancement", International Journal of Application or Innovation in Engineering & Management, Vol. 2, Issue 1.

[14] M. Ameur., A. Bouhjar, F. Boukhris, A. Boukouss, A. Boumalk, M. Elmedlaoui, E. Iazzi, and H. Souifi, "Initiation à la langue Amazighe", Publications de l'IRCAM, 2004.

[15] Fatima Amounas, 2015. "Enhanced Elliptic Curve Encryption Approach of Amazigh alphabet with Braille representation", International journal of Computer Science & Network Solutions, Vol. 3.No. 8, pp. 1-9.

[16] M. Ameur, A. Bouhjar, F. Boukhris, A. Boukouss, A. Boumalk, M. Elmedlaoui, and E. Iazzi, "Graphie et orthographe de l'Amazighe", Publications de l'IRCAM, 2006.

[17] F. Amounas and E.H. El Kinani, 2012. "Cryptography with Elliptic Curve using Tifinagh Characters", Journal of Mathematics and System Science 2, pp.1-6.

[18] F. Amounas and E.H. El Kinani, 2012. "Fast Mapping Method based on Matrix approach For Elliptic Curve Cryptography", International Journal of Information & Network Security, Vol.1, No.2, pp. 54-59.

[19] Geetha G and Padmaja Jain, 2014. "Implementation of Matrix based Mapping Method Using Elliptic Curve Cryptography", International Journal of Computer Applications Technology and Research, Vol. 3, Issue 5, pp. 312-317.

[20] Priti V. Bhagat, Kaustubh S. Satpute and Vikas R. Palekar, 2013. "Reverse Encryption Algorithm: A Technique for Encryption & Decryption", International Journal of Latest Trends in Engineering and Technology, Vol. 2 Issue 1.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 9, 584 - 589, 2016, ISSN:- 2319–8656

New Enhanced Method for Managing Database Project via Genetic Algorithm And Fuzzy Logic

Maryam KaivanluShahrestanaki
Department of Computer, Ferdows Branch,
Islamic Azad University
Ferdows, Iran

Dr.Hooman Kashanian
Department of Computer ,Ferdows Branch,
Islamic Azad University
Ferdows, Iran

Abstract: Practicable establishment between Project and its progress procedure is vital to yield the finest outcome but uncertainty is great challenge in such kind of formation. Set of related doings that has varying uncertainty characteristic named as Project. To manage such varying uncertain situation, it is vital to mold progress procedure flexibly as per terms and circumstances. This article has made deep endeavor for revision the phases to be commenced to design and discovers flexible methodology. This method is not combination or hybrid association of accessible methodology. It is providing independence to progress expert for select the progress method as per development needs of projects contributory module for project expansion. Main idea of this paper is to rectify the present difficulties and troubles in improvement process by representing 'Flexible approach' for development on the origin of 'Selection of software expansion procedure suggested system (SOSDMAS)-fuzzy expert system' for making competent development with possible and need base formation

Key words: Software design, Revolutionary algorithms, Software development, Genetic algorithm

1. INTRODUCTION

A Data acquisition (DAQ) system in common involves of The basic notion of the software expansion productivity depends mainly on two aspects, one is 'project' and another is 'development methodology' [24]. Thus, logically the success of project development process depends on the establishment of project and its development procedure. The revision reveals that, every project is unique but has set of contributory module with variable uncertainty level, on other hand it has effective option (development methodology) with best practice [14,15]. Therefore, it is challenging to define that which development procedure is appropriate for efficient development that will give best result, since each one has difficulties. Such kind of environments escalations the level of uncertainty and intricacy. In such condition it is very precarious and crucial to formulate possible formation of project and its development methodology as per its best practice for efficient development [19].Often it is mentioned as the "black art" or "brain tester" [11,22]. Now, development specialist do not have a magic wand threw that they formulate alignment between the project's interacting mechanisms and development procedure and project success.

Usually, 'success can be found by either by luck or insight of decreasing failure'. It is not only philosophy but our personal experience that in such inconstant variable circumstance, our method becomes flexible as per need of terms and situation to reduce failure. With this stimulus to rectify the failure part, this revision is carried out. The scientists have made workaholic ardent efforts by the way of distinctive permutations and combinations variables in the technological factors and analysis of relevant literature on "software development productivity and investigate its impact factors to reformulate and customize software and its strategy to gain expectation and design fallouts on the foundation of flexible approach.

In the paper planned in V sections:

- Section I deals with overview of title of this paper.
- Section II deals with limits in Software Development Productivity, treats and chance in software development process
- Section III deals with procedure and observation.
- Section IV is lead with debate supported by implication, interpretation and recommendation.
- The result and future work of study is depicted in section V.

SECTION II:

2.1 Restrains in Software Expansion Productivity:

Many literate writers study explored that uncertainty is seed of threats in feasible formation of Project and its development procedure in software development process. Because of rigidity, plan driven method is incompetent to fetch such uncertain milieu [16]. Practice driven method provides competent result in such environment [8]. But still software expansion industries fetch the problem of failures or overruns [9, 18]. Further though we have mix or hybrid method which treated one expansion methodology [13, 20 and 23]. In this review 'how one development procedure fits for all contributory modules development'?

2.2 Challenges for Software Expansion Procedure:

To tackle the uncertainty is big challenge in front of expansion procedure. However it has practice driven method based particularly adopt uncertainty [15, 23] but project is set of contributory module that has variable aspect and level of uncertainty. So rather than variation of uncertainty development specialist must fetch the uncertainty by recognizing its sources and level. But projects contributory modules subsequent formation returns uncertainty module to module. In such condition there may be risks of an inaccurate handling of design method or hasty decision, which may be prove to be wrong later [2, 3]. As consequence, expansion process become uncomfortable with variance situation and rise challenging state or overruns. Later it is one of the robust causes and limit for trip down the software project.

In such an uncertain condition, precious question is "How development specialist could formulate the project and its development procedure on the basis of feasible consideration?"

2.3 Chance for Software Development Procedure:

Based on review reports, there are some success stories. In such an uncertain condition 'How they become the successes or 'Do they have any mystic wand?' The answer is undeniably no because formulating the project and its expansion methodology on the basis of feasible consideration is very straight forward once one understands level of indecision and its sources [11, 22].This revision proposes the development practitioner comprehend project with its parallel activation formation because it distinguishes uncertainty its root sources module which is very helpful for allocation need base development approach as per to gain best productivity on the basis of flexibility in the sense of 'Just Utilization Gaining Approach And Deploy'.

Therefore our theory is "Flexible method for software expansion transfers opportunity for feasible formation of project and its development methodology"

SECTION –III: PROCEDURE

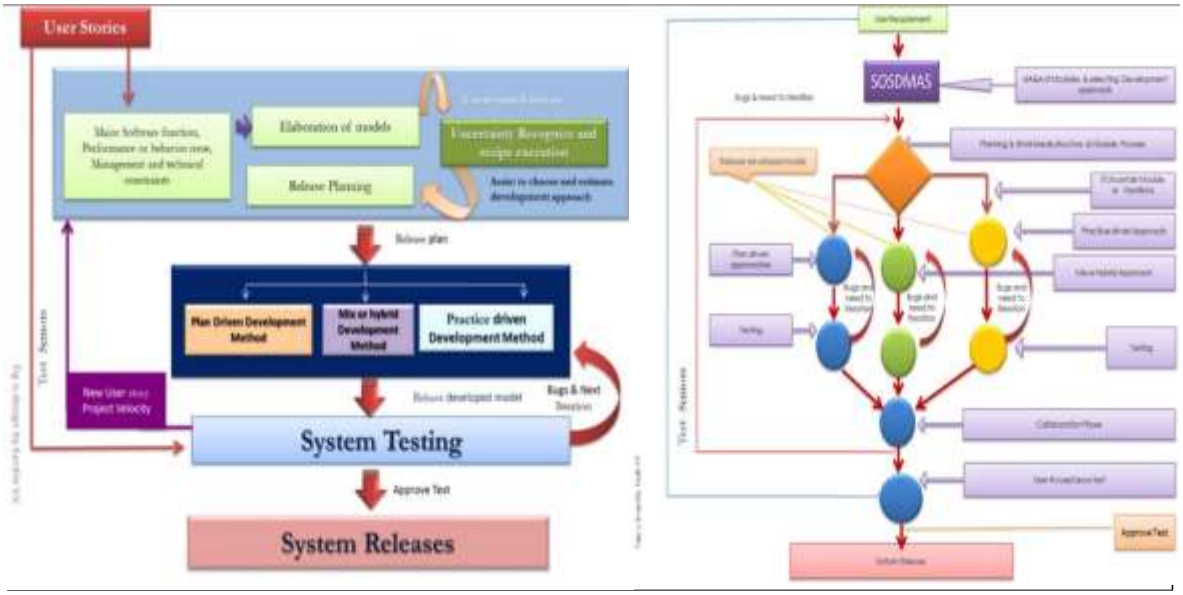
3.1 Qualitative Solution for theory: *In this section the* procedure of the revision has been discussed. The inductive method with qualitative solution is exploited to test the estimated hypothesis by experimental case study.

3.2 Implementation of 'Flexible method' in the software expansion process.

Flexible method: As the flexible methods principle goals to provide choice to decision maker to adopt the change as per terms and condition in accessible situation for produce better performance. It can be employed with such capability to deal with both predictable and unpredictable changes [1, 10, 12, and 17]. This method is adapting those ways for solution which is appropriate for accurate formulation of expansion process with feasible consideration.

It is appropriate for implementation in expansion process for development specialist. So, at the preliminary stage when user stories cognize at that time expansion practitioner can easily recognize uncertainty, its root sources with parallel formation contributory modules by rule base fuzzy expert system before modeling process for allocating suitable development methodology module wise for development. On the foundation of CMMI Level-2 & 3,
"To make sure that accessible processes are maintained at the time of stress, organization use causal analysis to identify and resolve the issues affecting performance and promote the dissemination of finest practices by Organizational Process Performance, Quantitative Work Management".

Table 1: Module wise Uncertainty level and its allocated development methodology by SOSDMAS



3.3Illustrative investigational case study using Flexible methodology

For the determination of validation and to test the flexible model we exploit system testing and act testing on one experimental case study, let's consider the 'Log in' form as a project. User specifies necessities of 'Log in' form module application that generates and preserves user's authentication of administration. It accumulates password from the user and authenticate it. It lets the new user to add information details. Give ability to recover the forgotten password by one or two phase of verification. Like subsequent screen as a project. This module has three sub actions as: Login ability, Sign up and forgot password.

Figure 1: Illustrative Project need screen.

With own feature of project and necessity, development practitioner can surly celebrate this module along with practice driven method with outward appearance of agile development methodologies instead of plan driven approach. Because of its sub activity it enable recover the forgotten password by one or two step verification. 'One phase verification' celebrates some hint question or substitute mailing address. In two phase authentication it will require alternative service like mobile no, or it can regenerate randomly and received by the head of department. It is not approve. But when we focus on remaining two sub activities evidently it is notable in certain aspect.

But when we apply 'Selection of software expansion methodology advisory system' on the above case study the outcome interprets that above case study celebrates intricate aspect. So with this concern, researcher assigns three teams on same platform (core java) with plan, practice driven and flexible driven method following team as per following table for validate estimated theory.

These three expansion team included 4computer science topper student and 2 working software professional with 2 year experience in well apparent software development firm for development purpose of above case study.

Here we assign team on same platform (core java) with owed development approach as per following table.

Table 2. formation of expansion team for act testing of experimental case study development

Team No	Team members	Owed development approach
I	2 software professional	Practice driven (SCURM)

II	1 M.Sc. [Computer], 1 U.G Level	Flexible method
III	1 M.C.A. [Computer], 1 U.G Level	Plan driven method (Water fall approach)

Result given by above team for develop experimental case study:

Table 3: Effort arrangement in Hours on Each Stage for above experimental case study

Strategy Driven	*Time (Hrs.)*	*Practice Driven*	*Time (Hrs.)*	*Flexible Method*	*Time (Hrs.)*
Initial Investigat	*5*	*Inceptio*	*5*	*Inception*	*5*
Necessity	*7*	*Elaborat*	*2*	*Elaborati*	*3*
System E	*5*	*Construc*	*7*	*Construc*	*7*
System D	*10*	*Testing &*	*15*	*Testing &*	*10*
System Testing	*8*				
System Execution	*2*				
Total	*37*		*29*		*25*

Section – IV Observation and assessment

4.1 Statement

Result discovers that flexible method driven development team is substantial because it takes less development time as compared to plan driven and practice driven development teams for above experimental case study development.

We are detected that,

< Plan driven method following team activate systematically but it wonders behind alteration In contras

Practice driven approach following team effectively bid on project development but it is irritated by its expert perception.

< The teams following Flexible method succeed to make it competent or need base development but it demands high coordination between team members.

4.2 Assessment

With this respect, as per our laboratory experiment the planned flexible model is efficient to reduce development time and minimizes risks. This will allow the software expansion to keep it low with suitable quality but it needs high monitoring and coordination in development process.

Additionally less expansion time reduces stress of development team and reduces development cost. It is directly interlinked with resource and schedule of expansion process. It also helps to make competent development or quality works that surly improve productivity.

Above thoughtfulness directly indicate "Flexible method for software development transfers chance for feasible formation of project and its development methodology"

4.3 Recommendation: It is true that we have a unlimited tool in the case of practice driven method that it adopt the uncertainty on the basis of agility. It delivers ability to expansion process to rapidly change its stage and direction in a particular way but the problem of failures or overruns remain same. Moreover we have newly arrived notion of mix or hybrid approach. Nevertheless it has mixture of both approach but it is treated as one type procedure. Here we never miscalculate the expansion process naturally involves inconstant changeable environment and one typed solution is not suitable to it. This revision reveals that problem is not sited in expansion methodology. It is located in formulation of 'Project', 'its expansion methodology' and 'success'. It should be conceivable by flexible formation or couple the project with finest practice of software development methodology on the basis of need to handle the uncertainty rather than adopting it. It should offer ability to choose the need base formation by taking into account the certainty or uncertainty aspect distinctly. But, in that problem, there is not much care reported or received so far.

4.4 Carrying out: Every software expansion organization have own procedure for project development management. It may be vary from the

organization to organization but it needs to address similar problems contain for development process. Implementation of flexible method is convenient for expansion practitioner for development. It needs carrying out only uncertainty distinguishes and recipe process with projects parallel activation formation for clarification and validation of contributory modules level of uncertainty before modelling procedure. Through that expansion specialist be able assign project's needing base development approach as per its best practice for making efficient development. Left over process remains as it is.

Figure 2: Performance of flexible method.

Section –V

5.1 Result:

In this section this revision released that possible formation of project and its expansion methodology as per its best practice formulation is not crucial or critical. Often it is denoted as the "black art" or "brain tester". It comes in presence with magic wand in the sense of flexible method. This revision hopefully conveys to the expansion specialist and trust that flexible approach transfers opportunity to enhance software development productivity. A flexible method may not be a complete solution for all software expansion difficulties; it could be opposed for those organizations that follow up one software expansion methodology to execute projects. But there is a demand to change as per terms and circumstances in available situation for enhancement.

5.2 Opportunity of further work:

As per literature analysis it is yet to publish announcement, which can explores the policy for decreasing the ratio of project challenges and cost overrun /time element. This revision found that, there is chance for lessening the ratio of project challenging and cost overrun /time element with flexible approach. In this section the revision learner agreed that this is very preliminary part of practice. There must be comprehensive authentications for proposed method.

REFERENCEs

1. A. M. Gavkare, N. L.Nanaware, A. R.Waghmare, G.B. Taware, A. D.Surdi (2012)"Study of Flexibility, Agility and Reaction Time in Circus Artists" International Journal of Recent Trends in Science And Technology, E-ISSN 2239-8109, Volume 1, Issue 2, 2011 pp 49-55
2. Barry W. Boehm, TRW(2001)Improving Software productivity (csse.usc.edu/csse/TECHRPTS/1988/usccse87-502/usccse87-501.pdf last access 31/02/12)
3. Boehm, B., & Turner, R. (2005). Balancing Agility and Discipline: A guide for the Perplexed. Addison- Wesley.(book)
4. ChitraG.Desia ,Kardile V V(2014)"uncertainty is not dilemma in software project development process" IEEE ISBN:978-1-7695-4437-3
5. ChitraG.Desia ,Kardile V V(2014)"uncertainty is not dilemma in software project development process" IEEE ISBN:978-0-7695-4437-2
6. ChitraG.DesiaKardile V.V.(2014)"Jugaad"-the creativeness For Selection of software development methodology advisory system- Fuzzy expert system" presented at 2nd International Conference on Computer and Communication Technologies (IC3T 2014) and published by Springer-Verlag
7. ChitraG.DesiaKardile V.V.(2014) Challenges in software development productivity –A literature review" Advances in computational research ISSN : 0975 –3973 DIO:10.97351975 – 3273 Volume 7,Issue 1
8. Chow, T., & Cao, D. (2009). A Survey of Critical Success Factors in Agile Software Projects. The Journal of Systems and Software, 81 (6), 961-972.
9. Dr. Kevin Thompson, (2012)agile journal productivity report www.agilejournal.com › Articles › Columns › Articles)
10. FerielDaoudiSelminNurcan(2009) "A benchmarking framework for methods to design flexible business processes"
11. Jens Christian Refsgaard a,*, Jeroen P. van der Sluijs b, Anker LajerHøjberg a, Peter A. Vanrolleghemc,dJ.C. Refsgaard et al. (2009) Uncertainty in the environmental modelling process e A framework and guidance Environmental Modelling & Software 29 (2007) 1543-1556
12. Johan Bekar (1999) "Agility and flexibility what is difference" cranfield school of management ISBN 1 85905 068 3
13. Juyun Cho reported "A hybrid software development method for large-scale projects: rational unified process with

scrum" lume X, No. 2, 2008 341 Issues in Information Systems

14. Lemétayer, J. (2012). Identifing the Critical Success Factors in Project Management Methodology Fit. PMI Global Congress Proceedings. Melbourne, Australia.

15. Little, T. (2008). Context-Adaptive Agility: Managing Complexity and Uncertainty. IEEE Software, 22 (3), 28-3 6.

16. M. A. Awad (2008) "A Comparison between Agile and Traditional Software Development Methodologies"

17. Prestone G Smith, Jeff Oltmann (2015) "flexible project management: extended the agile technique beyond software project" PMI Global Congress Processing-Washington DC 2015.

18. Scott W. Ambler (2015) Defining Success, by. Dr. Dobb's Journal. source :2014 IT project success survey,5www.ambaysoft.com/surveys/ success2015.html

19. Shenhar, A. J. (2007). One Size Does Not Fit All Projects: Exploring Classical Contingency Domains. Management Science, 47 (3), 394-412

20. SiddharthSharadChandak.and Vishnu Rangarajan.is a Project Manager at Cognizant "Flexibility in Software Development Methodologies: Needs and Benefits"cognizant 20-20 insights , november 2011(last access 25 march 2013at google search)

21. Walker Royse(1999) "Software project management : Unified framework" forwarded by Barry Bohem The Addison-Wesley object technology series

22. Walker, W.E., Harremoe s, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., 2008. Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 7 (1), 5e16

23. William Chaves de Souza Carvalho, Pedro Frosi Rosa, Michel dos Santos Soares reported (2018)" A hybrid approach to integrate agile and traditional software development processes "

24. Wysocki, R. R. (2008). Efective Project Management: Traditional, Agile, Extreme (6th ed.). Indianapolis, IN: Wiley.

International Journal of Computer Applications Technology and Research
Volume 5–Issue 9, 590-594, 2016, ISSN:-2319–8656

Randić Index of Some Class of Trees with an Algorithm

H. S. Ramane
Department of mathematics,
Karnatak University,
Dharwad, India.

R. B. Jummannaver
Department of mathematics,
Karnatak University,
Dharwad, India.

V. K. Kyalkonda
Department of stastistics,
Karnatak University,
Dharwad, India.

Abstract: The Randić index $R(G)$ of a graph G is defined as the sum of the weights $(d_G(u)d_G(v))^{-1/2}$ over all edges $e = uv$ of G. In this paper we have obtained the Randić index of some class of trees and of their complements. Also further developed an algorithmic technique to find Randić index of a graph.
Keywords: Algorithm, Degree of a vertex, Randić index, Tree.

1. INTRODUCTION

Let G be an undirected graph without loops and multiple edges with n vertices and m edges. Let $V(G) = \{v_1, v_2, \dots, v_n\}$ be the vertex set of G and $E(G) = \{e_1, e_2, \dots, e_n\}$ be the edge set of G. There are many types of indices, some based on distance of a graph and some other based on degrees of vertices of graphs. In 1975, the Randić index was proposed by the chemist Milan Randić [6] under the name ''branching index''. The Randić index $R(G)$ of a graph G is defined as the sum of the weights $(d_G(u)d_G(v))^{-1/2}$ over all edges $e = uv$ of G, where $d_G(u)$ is the degree of a vertex u in G. That is,

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u)d_G(v)}}$$

I. Gutman, et., al studied its mathematical properties and summarized in recent books [1,2]. The history of this index is described in [7, 8].It has been found that the Randić index correlates well with the harmonic index [4]. The expressions for the harmonic index and Randić index of the generalized transformation graphs and for their complement graphs were obtained in [5].The adjacency matrix of a graph G is the $n \times n$ matrix $A(G) = [a_{ij}]$, in which $a_{ij} = 1$ if v_i is adjacent to v_j and $a_{ij} = 0$, otherwise [3]. The harmonic index of some trees are obtained and an algorithm for the evaluation of the index is developed in [6].

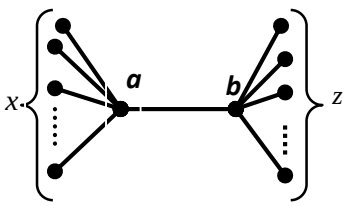


Fig.1: T_1

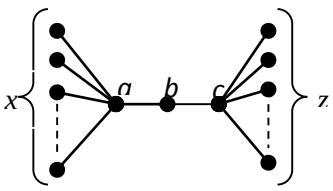


Fig.2: T_2

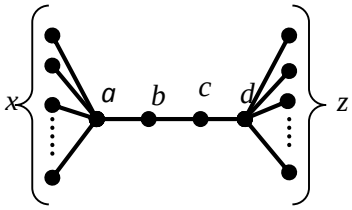


Fig3: T_3

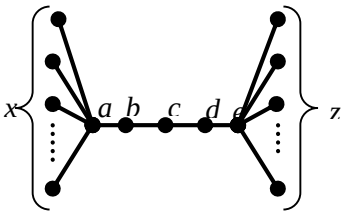


Fig.4: T_4

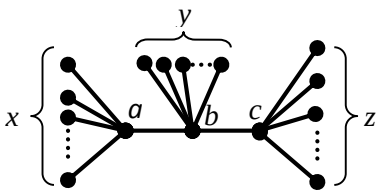


Fig.5: T_5

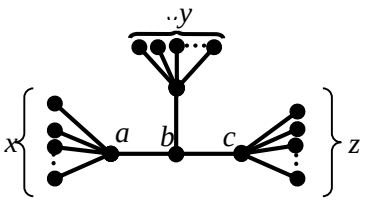


Fig.6: T_6

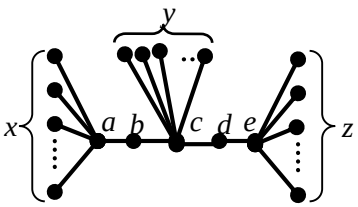


Fig.7: T_7

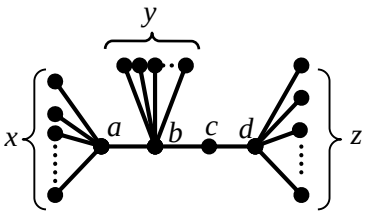


Fig.8: T_8

2. RESULTS:

Proposition 2.1: If T_1 is a tree with n vertices as shown in *Fig.1*, then the Randić index of T_1 is

$$R(T_1) = \frac{x}{\sqrt{x+1}} + \frac{n-x-2}{\sqrt{n-x-1}} + \frac{1}{\sqrt{(x+1)(n-x-1)}}.$$

Proof: Without loss of generality, consider the vertices a, b as shown in *Fig. 1*, where $d_{T_1}(a)=x+1$, $d_{T_1}(b)=n-x-1$. Partition the edge set $E(T_1)$ into 3 sets E_1, E_2 and E_3 such that $E_1=\{uv \mid d_{T_1}(u)=1 \text{ and } d_{T_1}(v)=x+1\}$, $E_2=\{uv \mid d_{T_1}(u)=1 \text{ and } d_{T_1}(v)=z+1\}$, $E_3=\{ab\}$. It is easy to see that $|E_1|=x$, $|E_2|=n-x-2$, $|E_3|=1$.

Therefore,

$$R(T_1) = \sum_{uv \in E(T_1)} \frac{1}{\sqrt{d_{T_1}(u)d_{T_1}(v)}}$$

$$= \sum_{uv\in E_1(T_1)} \frac{1}{\sqrt{d_{T_1}(u)d_{T_1}(v)}} + \sum_{uv\in E_2(T_1)} \frac{1}{\sqrt{d_{T_1}(u)d_{T_1}(v)}} + \sum_{uv\in E_3(T_1)} \frac{1}{\sqrt{d_{T_1}(u)d_{T_1}(v)}}$$

$$= \sum_{uv\in E_1(T_1)} \frac{1}{\sqrt{(x+1)(1)}} + \sum_{uv\in E_2(T_1)} \frac{1}{\sqrt{(n-x-2+1)(1)}} + \sum_{uv\in E_3(T_1)} \frac{1}{\sqrt{(x+1)(n-x-2+1)}}$$

$$= x\frac{1}{\sqrt{(x+1)}} + (n-x-2)\frac{1}{\sqrt{(n-x-2+1)}} + \frac{1}{\sqrt{(x+1)(n-x-2+1)}}$$

$$R(T_1) = \frac{x}{\sqrt{x+1}} + \frac{n-x-2}{\sqrt{n-x-1}} + \frac{1}{\sqrt{(x+1)(n-x-1)}}.$$

The following proposition 2.2 can be proved in analogous to the proposition 2.1.

Proposition 2.2: If T_i is a tree with n vertices as shown in *Fig. i*, then the Randić index of T_i, $i=2,3,4$ as follows,

$$(i)\ R(T_2) = \frac{x}{\sqrt{x+1}} + \frac{n-x-3}{\sqrt{n-x-2}} + \frac{1}{\sqrt{2(x+1)}} + \frac{1}{\sqrt{2(n-x-2)}}$$

$$(ii)\ R(T_3) = \frac{1}{2} + \frac{x}{\sqrt{x+1}} + \frac{n-x-4}{\sqrt{n-x-3}} + \frac{1}{\sqrt{2(x+1)}} + \frac{1}{\sqrt{2(n-x-3)}}$$

$$(iii)\ R(T_4) = 1 + \frac{x}{\sqrt{x+1}} + \frac{n-x-5}{\sqrt{n-x-4}} + \frac{1}{\sqrt{2(x+1)}} + \frac{1}{\sqrt{2(n-x-4)}}$$

Proposition 2.3: If T_5 is a tree with n vertices as shown in *Fig.5*, then the Randić index of T_5 is

$$R(T_5) = \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+2}} + \frac{n-x-y-3}{\sqrt{n-x-y-2}} + \frac{1}{\sqrt{(x+1)(y+2)}} + \frac{1}{\sqrt{(y+2)(n-x-y-2)}}.$$

Proof: Without loss of generality consider the vertices a,b,c as shown in *Fig. 5*, where $d_{T_5}(a)=x+1$, $d_{T_5}(b)=y+2$, $d_{T_5}(c)=z+1$. Partition $E(T_5)$into 5 sets E_1, E_2, E_3, E_4, and E_5such that $E_1=\{uv \mid d_{T_5}(u)=1$ *and* $d_{T_5}(v)=x+1\}$, $E_2=\{uv \mid d_{T_5}(u)=1$ *and* $d_{T_5}(v)=y+2\}$, $E_3=\{uv \mid d_{T_5}(u)=1$ *and* $d_{T_5}(v)=z+1\}$, $E_4=\{ab\}$, $E_5=\{bc\}$. It is easy to see that $|E_1|=x$, $|E_2|=y$, $|E_3|=z$, $|E_4|=|E_5|=1$.

Therefore,

$$R(T_5) = \sum_{uv\in E(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}}$$

$$= \sum_{uv\in E_1(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}} + \sum_{uv\in E_2(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}} + \sum_{uv\in E_3(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}} + \sum_{uv\in E_4(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}} + \sum_{uv\in E_5(T_5)} \frac{1}{\sqrt{d_G(u)d_G(v)}}$$

$$= \sum_{uv\in E_1(T_5)} \frac{1}{\sqrt{(x+1)(1)}} + \sum_{uv\in E_2(T_5)} \frac{1}{\sqrt{(y+2)(1)}} + \sum_{uv\in E_3(T_5)} \frac{1}{\sqrt{(z+1)(1)}} + + \sum_{uv\in E_4(T_5)} \frac{1}{\sqrt{(x+1)(y+2)}} + \sum_{uv\in E_5(T_5)} \frac{1}{\sqrt{(y+2)(z+1)}}$$

$$= \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+2}} + \frac{z}{\sqrt{z+1}} + \frac{1}{\sqrt{(x+1)(y+2)}} + \frac{1}{\sqrt{(y+2)(z+1)}}$$

Here we have $n=x+y+z+3$. By replacing $z=x+y+z+3$, the above equation reduces to

$$R(T_5) = \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+2}} + \frac{n-x-y-3}{\sqrt{n-x-y-2}} + \frac{1}{\sqrt{(x+1)(y+2)}} + \frac{1}{\sqrt{(y+2)(n-x-y-2)}}.$$

Proposition 2.4: If T_i is a tree with n vertices as shown in *Fig. i*, then the Randić index of T_i, $i=6,7,8$ is as follows,

$$R(T_6) = \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+1}} + \frac{n-x-y-4}{\sqrt{n-x-y-3}} + \frac{1}{\sqrt{3(x+1)}} + \frac{1}{\sqrt{3(y+1)}} + \frac{1}{\sqrt{3(n-x-y-3)}}$$

$$R(T_7) = \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+2}} + \frac{n-x-y-5}{\sqrt{n-x-y-4}} + \frac{1}{\sqrt{2(x+1)}} + \frac{2}{\sqrt{2(y+2)}} + \frac{1}{\sqrt{2(n-x-y-4)}}$$

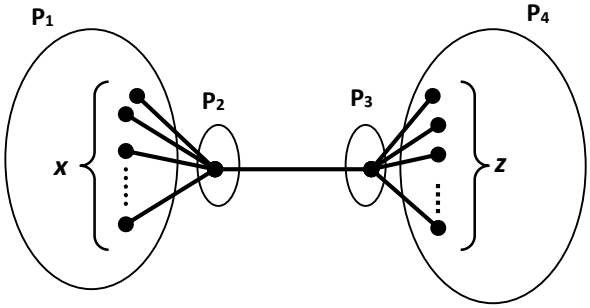
$$R(T_8) = \frac{x}{\sqrt{x+1}} + \frac{y}{\sqrt{y+2}} + \frac{n-x-y-4}{\sqrt{n-x-y-3}} + \frac{1}{\sqrt{(x+1)(y+2)}} + \frac{1}{\sqrt{2(y+2)}} + \frac{1}{\sqrt{2(n-x-y-3)}}$$

3. RESULTS FOR COMPLEMENTS

The complement of a graph G, denoted by $\bar{G}$ is a graph with vertex set $V(G)$ and two vertices in $\bar{G}$ are adjacent if and only if they are not adjacent in G [3].

Preposition 2.5: If $G=T_1$ is a tree with n vertices and m edges as shown in *Fig.1*, then the Randić index of complement of $T_1$is

$$R(\bar{T}_1) = R(\bar{G}) = \frac{n-3}{2} + \frac{n-x-2}{\sqrt{(n-2)(n-x-2)}} + \frac{x}{\sqrt{x(n-2)}}$$



Proof: Consider the partition $\mathbf{P_1}$, $\mathbf{P_2}$, $\mathbf{P_3}$ and $\mathbf{P_4}$ ofvertex set of T_1as shown in *Fig.9* . Easily we can note that $|\boldsymbol{P_1}|=x$, $|\boldsymbol{P_2}|=1$, $|\boldsymbol{P_3}|=1$, $|\boldsymbol{P_4}|=n-x-2$. Therefore,

$$R(\bar{T}_1) = \sum_{uv\in E(\bar{T}_1)} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}}$$

$$= \sum_{uv\in(P_1)} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}} + \sum_{\substack{u\in(P_1)\\ v\in(P_3)}} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}} +$$

$$\sum_{\substack{u\in(P_1)\\ v\in(P_4)}} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}} + \sum_{\substack{u\in(P_2)\\ v\in(P_4)}} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}} + \sum_{uv\in(P_4)} \frac{1}{\sqrt{d_{\bar{G}}(u)d_{\bar{G}}(v)}}$$

$$= \sum_{uv\in(P_1)} \frac{1}{\sqrt{(n-1-1)(n-1-1)}} + \sum_{\substack{u\in(P_1)\\ v\in(P_3)}} \frac{1}{\sqrt{(n-1-1)(n-1-(n-x-1))}}$$

$$+ \sum_{\substack{u\in(P_1)\\ v\in(P_4)}} \frac{1}{\sqrt{(n-1-1)(n-1-1)}} + \sum_{\substack{u\in(P_2)\\ v\in(P_4)}} \frac{1}{\sqrt{(n-1-(x+1))(n-1-1)}}$$

$$+ \sum_{uv\in(P_4)} \frac{1}{\sqrt{(n-1-1)(n-1-1)}}$$

$$= \sum_{uv\in(P_1)} \frac{1}{(n-2)} + \sum_{\substack{u\in(P_1)\\ v\in(P_3)}} \frac{1}{\sqrt{(n-2)x}} + \sum_{\substack{u\in(P_1)\\ v\in(P_4)}} \frac{1}{(n-2)}$$

$$+ \sum_{\substack{u\in(P_2)\\ v\in(P_4)}} \frac{1}{\sqrt{(n-x-2)(n-2)}} + \sum_{uv\in(P_4)} \frac{1}{(n-2)}$$

$$= \frac{n(n-1)}{2(n-2)} + \frac{x}{\sqrt{x(n-2)}} + \frac{x(n-x-2)}{(n-2)} + \frac{n-x-2}{\sqrt{(n-x-2)(n-2)}}$$

$$+ \frac{(n-x-2)(n-x-3)}{(n-2)}$$

$$R(\bar{T}_1) = \frac{n-3}{2} + \frac{n-x-2}{\sqrt{(n-2)(n-x-2)}} + \frac{x}{\sqrt{x(n-2)}}$$

Proposition 2.6: If T_i is a tree with n vertices as shown in *Fig. i*, $i=2,3,4,5$ then the Randić index of complement of T_i is as follows,

$$R(\bar{T}_2) = \frac{n^2-7n+12}{2(n-2)} + \frac{x}{\sqrt{(n-2)(n-3)}} + \frac{x}{\sqrt{(n-2)(x-3)}} + \frac{x}{\sqrt{(n-2)(x+1)}}$$

$$+ \frac{1}{\sqrt{(n-x-2)(x+1)}} + \frac{n-x-3}{\sqrt{(n-x-2)(n-2)}} + \frac{n-x-3}{\sqrt{(n-3)(n-2)}}$$

$$R(\bar{T}_3) = \frac{n^2-9n+20}{2(n-2)} + \frac{2x}{\sqrt{(n-2)(n-3)}} + \frac{x}{\sqrt{(n-2)(x+2)}}$$

$$+ \frac{1}{\sqrt{(n-x-2)(n-3)}} + \frac{1}{\sqrt{(n-x-2)(x+2)}} + \frac{n-x-4}{\sqrt{(n-x-2)(n-2)}}$$

$$+ \frac{1}{\sqrt{(n-3)(x+2)}} + \frac{2(n-x-4)}{\sqrt{(n-3)(n-2)}}.$$

$$R(\bar{T}_4) = \frac{n^2-11n+30}{2(n-2)} + \frac{3x}{\sqrt{(n-2)(n-3)}} + \frac{x}{\sqrt{(n-2)(x+3)}}$$

$$+ \frac{2}{\sqrt{(n-x-2)(n-3)}} + \frac{1}{\sqrt{(n-x-2)(x+3)}} + \frac{n-x-5}{\sqrt{(n-x-2)(n-2)}}$$

$$+ \frac{1}{n-3} + \frac{2}{\sqrt{(n-3)(x+3)}} + \frac{3(n-x-5)}{\sqrt{(n-3)(n-2)}}.$$

$$R(\bar{T}_5) = \frac{x(x-1)}{2(n-2)} + \frac{x}{\sqrt{(n-2)(n-z-2)}} + \frac{x}{\sqrt{(n-2)(n-y-3)}} + \frac{xy}{n-2}$$

$$+ \frac{xz}{n-2} + \frac{y}{\sqrt{(n-x-2)(n-2)}} + \frac{1}{\sqrt{(n-x-2)(n-z-2)}} + \frac{z}{\sqrt{(n-x-2)(n-2)}}$$

$$+ \frac{y(y-1)}{2(n-2)} + \frac{y}{\sqrt{(n-z-2)(n-2)}} + \frac{yz}{n-2} + \frac{z}{\sqrt{(n-y-3)(n-2)}} + \frac{z(z-1)}{2(n-2)}$$

4. ALGORITHM:

➢ An algorithm to find the Randić index of a graph.

Step 1: START

Step 2: Declare: *a*[25][25],*d*[25],m as integers

sum1,*s*[25],sum, *ts=0* as floating points.

Step 3: Read *m*, *a*[*i*][*j*].

Step 4: Compute : Degree of each vertex of given graph

for *i* to *n*

d[*i*] ← *0*

for *j* to *n*

d[*i*] ← *d*[*i*] + *a*[*i*][*j*]

Display: Degree *d*[*i*] of vertex *i*

Step 5: Check the condition, if *a*[*i*][*j*]=*1* is true

Display: Vertex *i* is adjacent to vertex *j*

Step 6 : Multiply corresponding degrees of adjacent vertices

sum ← *d*[*i*]**d*[*j*]

Step 6: Display the sum of multiples of adjacent vertices degree

$ts \leftarrow ts+(1/\sqrt{sum})$

Step 7: Display the Randić index by dividing total sum *ts* by 2.

Step 8:STOP

Illustartion:

G:

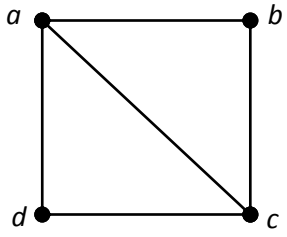


Fig. 10

We represent the graph G by adjacency matrix,

$$A(G) = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

In this matrix *a, b, c, d* represents the vertices of graph *G*. The element *1* in *A(G)* represents the adjacency between the vertices and *0* represents the non-adjacency between the vertices. Addition of elements of each row gives the degree of a corresponding vertex in *G*, ie., we get 3 by adding all the elements of a first row of adjacency matrix *A(G)*which is degree of vertex '*a'* in graph G. similarly we get other vertex degrees by adding corresponding row. Using this we calculate degree of each vertex and store it in *d[i]* by using for loop.

The outer loop iterates *i* times and the inner loop iterates *j* times, the statements inside the inner loop will be executed a total of *i*j* times. It is because, inner loop will iterate *j* times for each of the *i* iterations of the outer loop. This means the outer and inner loop are dependent on the problem size ie., here we considered size is *n*, the statement in the whole loop will be executed $O(n^2)$ times. In the loop int *i=0*, this will be executed only once. The time is actually calculated to *i=0* and not the declaration, *i<n* this will be executed *n+1* times, *i++* will be executed *n* times, *a[i][j]=1*, This will be executed *n* times (in worst case scenario).

As per the definition Randić index we multiply the degree of vertices which are adjacent, by adjacency matrix *A(G)*, we check the adjacency of one vertex to another by using if condition, then we multiply the degree of those adjacency vertices using *d[i]*,(This loop follows same procedure as explained for above loop so this also executed $O(n^2)$ times). Then we sum the multiplied valueof each adjacent vertices and each time we store the resulting value in one variable say *ts* as per the definition of Randić index,ie.,$1/\sqrt{sum}$. So for the above example we get final value of *ts* as 3.9266. We obtain the Randić index by dividing *ts* by 2.Therefore the Randić index of this example is 1.9633.

5. CONCLUSION

The results give explicit formulas for Randić index of certain class of trees and also for their complements. Further an algorithm with the help of adjacency matrix given to compute the Randić index of graph.

6. ACKNOWLEDGEMENTS

This work is supported by the University Grants Commission (UGC), Govt. of India for support through research grant under UPE FAR-II grant No. F14-3/2012 (NS/PE).

7. REFERENCES :

1. I. Gutman, B. Furtula (Eds.), Recent Results in the Theory of Randic´ Index,Uni. Kragujevac, Kragujevac, 2008.
2. X. Li, I. Gutman, Mathematical Aspects of Randic´-Type Molecular Structure Descriptors, Uni. Kragujevac, Kragujevac, 2006.
3. F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
4. X. Li, Y. Shi, A survey on the Randic index, MATCH Commun. Math. Comput. Chem., 59 (2008),127-156.
5. H. S. Ramane, B. Basavanagoud, R. B. Jummannaver, Harmonic index and Randić index of generalized transformation graphs, preprint.
6. H. S. Ramane, R. B. Jummannaver, S. C. Patil, "Harmonic index of some class of trees with an algorithm", IOSR Journal of Mathematics (IOSR-JM).12(2) (2016), 52-57
7. M. Randić, On characterization of molecular branching, J. Am. Chem. Soc. 97 (1975) 6609–6615.
8. M. Randić, On history of the Randi´c index and emerging hostility toward chemical graph theory.MATCH Commun. Math. Comput. Chem. 59(2008), 5–124 .
9. M. Randi´c, The connectivity index 25 years later. J. Mol. Graph. Model. 20(2001), 19–35.

International Journal of Computer Applications Technology and Research
Volume 5–Issue 9, 595-602, 2016, ISSN:-2319–8656

Factors inhibiting the adoption of ICT by Tamale Polytechnic lecturers for the training of the middle-level-manpower professionals in Ghana.

Zakaria Abukari
Lecturer, Dept. of Computer Science
Tamale Polytechnic
Tamale, Ghana

Abstract: Although the Ghanaian polytechnics have had computers and varied levels of ICT development for almost two decades now, ways to create effective IT-enabled teaching and learning methodologies have evolved slowly and patchily. This situation is gradually making the polytechnic trainees incompatible in the digital-enabled job markets. Coupled with this development is the fact that the internet has become the single and largest library and knowledge reservoir thus making it indispensable in the teaching and learning ambit. It has therefore become imperative and collective responsibility to identify the factors that inhibit the adoption of the technology by the tertiary teachers especially the Polytechnic Teachers Association of Ghana (POTAG) fraternity to bridge the digital gab to add more value to the polytechnic teachers and graduates and to raise their relevance in the industry. This research therefore comes in, with the case of the Tamale Polytechnic, to explore the challenges and recommend strategies to stakeholders. Descriptive survey methodology, which is capable of collecting background information and hard to find data without the researcher motivating or influencing respondents' responses, was used to arrive at our findings.

Keywords: ICT; Teaching and learning; Policy; LCD Projector; Sample size, Sample population

1. INTRODUCTION

Information and Communication Technology is emerging now as the pivot of change in every facet of life today and for this reason, it is widely embraced in the society today. Matthew(1997) states that the use of information and communication technology is clearly shaping the ways in which we learn, work and spend our leisure. As such, our success as individuals or as a nation depends on our ability to understand and use ICT. Today no one can afford to ignore the importance of computer technology in one's everyday life.

Teachers are expected to work paying more attention to students' autonomy and independence in the teaching and learning process. ICT is the major enabler in this direction. Much the same, the role of the teacher as a facilitator of the process should not be underestimated. In this era of digitization, it is time to question the old values of the traditional forms of education but on the other hand student-teacher collaboration can be mutually beneficial. Technology has created favourable conditions for teachers to help students to be engaged equitably as active partners in learning. Pedagogical communication is evidently important to renew the teacher-student communication style as new digital forms require new approaches not only on the process of teaching but also to the way of interacting. Pedagogical communication is known to be the interaction between a teacher and a student to achieve certain educational goals.

The quality of education depends, to a large extent, on the quality of teachers involved in its development and delivery. A qualified teacher will acknowledge the needs and interest of students, allowing them to learn at their own convenience, encourage learning, where necessary, intervene through remedial and enrichment instructions among others (MIE, 2004). ICT is considered a powerful tool for educational change and reform. Appropriate use of ICT can improve quality of education and connect learners to real-life situations (Lowther, et. al., 2008);(Weert and Tatnall, 2005). As Weert and Tatnall (2005) have stated, learning is a lifelong ongoing activity where learners change their expectations by seeking knowledge through ICT tools such as the internet and other electronic simulations that depart from the traditional approaches. Skills in using ICT will be an indispensable prerequisite for these learners.

Despite these outlined benefits of ICT in the teaching and learning environment, teachers in the tertiary institutions in Ghana are not yet harnessing these opportunities. It is against this backdrop that this research is attempting to investigate the underlying factors that demotivate the use of ICT in teaching and learning in the Tamale Polytechnic, an environment that does not differ much from the other nine (9) polytechnics in the country. Based on our findings, the appropriate recommendations shall be made to improve the situation. The fact remains intangible that the way forward for students' autonomy, effective teaching and learning is ICT.

2. STATEMENT OF THE PROBLEM

There have been exponential increases in the acquisition of computers in schools not just in the Western countries, but increasingly in developing ones as well. Although it is almost two decades, ways to use them effectively in many schools in the under-developed countries including Ghana have evolved slowly and patchily. The technological revolution has been beset by many factors that have kept the implementation of the educational

technology below the expectations in our societies. Whilst the Western countries, through IT-enabled teaching and learning systems, are gradually shifting from the traditional classrooms to the eClassrooms, developing countries are yet to take the advantage. In Ghana, there are adequate ICT facilities in the tertiary institutions such as computers, laboratories, high speed internet connections etc. yet many of the lecturers still deliver their lessons the traditional way through the use of blackboards (now whiteboards), lecture notes popularly called "handouts", manual assignments etc. One will question why teachers do not take advantage use of ICTs at their disposal to join the stream of change in teaching and learning. According to Gregoire et. al (1996), John and Sutherland (2004), the practical skills of the teacher and his/her attitude towards the use of ICT in teaching and learning contribute greatly to the students development. This research is therefore coming in to uncover the challenges, bottlenecks or call them inhibiting factors to the teachers' adoption of ICT in the tertiary institutions in Ghana using the Tamale Polytechnic as the case.

3. OBJECTIVES OF THE STUDY

This study seeks to explore the factors inhibiting polytechnic teachers' use of ICT in teaching and learning. To achieve our goal, the research addresses the following specific objectives:

- To find out the barriers hindering the integration of ICT into teaching and learning in Tamale Polytechnic.
- To assess the ICT infrastructural capacity for teaching and learning, in Tamale Polytechnic
- To determine the level of teacher's ICT knowledge and skills for teaching and learning in Tamale Polytechnic
- To examine the attitudes of teachers in the use of computers in Tamale Polytechnic.
- To recommend strategies for stakeholders to adopt for institutionalizing ICT use in teaching and learning.

4. SIGNIFICANCE OF THE STUDY

The impact of ICT on education is enormous. In the digital-enabled job markets, a graduate without IT competence is half-baked. The internet has become the single and largest library and knowledge reservoir making it therefore indispensable in the teaching and learning ambit. In this regard, identifying and addressing the factors that inhibit the adoption of the technology by the tertiary teachers especially the polytechnic teachers who are mandated to train the middle-level-manpower, will bridge the teaching digital gab adding more value to the polytechnic teachers and graduates making them more industry relevant. In specific terminologies, the importance of this research is summarized as:

i. The work will provide guidance to improve the use of ICTs in teaching and learning
ii. Teachers and graduates will be more industry compatible and relevant if the recommendations to the findings are implemented.
iii. This will also serve as a springboard for other researchers
iv. The research will showcase the impact of ICT in teaching and learning.

5. METHODOLOGY

Two principal research techniques were used. Formal and informal participatory methods were used to obtain the needed data for analyses. The informal methods involved observation and interviews of key stakeholders whilst the formal methods included the use of questionnaire for descriptive survey. According to Busha and Harter (1980) descriptive survey is capable of collecting background information and hard to find data and the researchers would not have the opportunity to motivate or influence respondents' responses. The technique is specially recommended for research where attitudes, ideas, comments and public views on a problem or issue are studied Sproull (1995), Iddrisu (2009).

A total of eighty (80) lecturers were randomly selected from ten (10) departments of the sixteen tertiary departments. The sample size was calculated according to Yamane (1973). The formula is provided below.

$$n = \frac{N}{1+N(e)^2}$$

where
n = is the required sample size. N= the population size
e = Tolerable error (which in this study was pegged at 0.05).
Proportional allocation was used to obtain the size that is supposed to be taken from each stratum (Table 1)

Table 1: Sample size

No	Name of Department	Population	Calculation	Sample size
1	Computer science	8	$8/100 \times 80$	6
2	Accountancy	16	$16/100 \times 80$	13
3	Sec. and mgt studies	11	$11/100 \times 80$	9
4	HCIM department	8	$8/100 \times 80$	6
5	Marketing	10	$10/100 \times 80$	8
6	Mechanical engineering	9	$9/100 \times 80$	7
7	Building Technology	10	$10/100 \times 80$	8
8	Agricultural engineering	11	$11/100 \times 80$	9
9	Industrial Arts	6	$6/100 \times 80$	5
10	Statistics department	11	$11/100 \times 80$	9
	Total	**100**		**80**

Both closed-ended and opened-ended questions were used in the questionnaire. The opened-ended questions were used to allow the respondents to express themselves without any given limit. The questionnaire was adapted and modified from Rodden (2008).
The data collected was checked for consistency. Statistical Package for the Social Sciences (SPSS) was used for the analysis.

6. PREVIOUS WORKS REVIEW

Reviewing previous works on related topic of research according to (Cisse, 2006) is a prudent approach to reducing repeated errors. Many researchers have looked at ICTs and education. Computers became common in the 1980's when personal computers became accessible to consumers. Since then there has been government policies that encouraged the mass production of computers for schools. Several researchers suggested that ICT is now an essential part of the education process (Pelgrum and Law, 2003); Hepp et. al (2004); Kozma (2008). Educational systems need to prepare students to adjust to and persist in this new technologically compelled society. This means preparing students for "ultimate learning in an information society" (Pelgrum and Law, 2003). In addition to this, early promoters of ICT integration into education saw it as a facilitator for change, nurturing skills in problem solving and critical thinking, as well as the development of student centred learning (McGrail, 2005).

According to Kozma (2008) there are three grounds for the introduction of ICT into education. The first one is the economic ground which refers to the role it can play in preparing students as future workers and in supporting financial development. The second is the social ground where ICT investment aims to elevate knowledge sharing, encourage cultural creativity, increase civic participation, make government services more accessible and finally enhance social cohesion. The third and final ground is the educational and pedagogic rationale where ICT can advance educational reform and advance educational management structures. Similarly, Hepp et. al (2004) broadly concur, finding three reasons for the use of ICT in education: the development of new skills for the information age, increased efficiency and the development of quality learning.

Whereas Kozma (2008) posits that there are three rationales for the introduction of ICT into education, Hawkridge (1990) proposes four rationales for the utilization of computers in schools. He notes these as social, vocational, pedagogical and catalytical. The social and vocational grounds point to the increased use of ICT in all spheres of human activity. The pedagogical and catalytically rationales relate to the effects of technology on students and schools. There are various views of others Bigum(1997); Hawkridge(1990); Drent and Meelissen (2008).

The use of ICTs improves all forms of information exchange, observation, learning and teaching. There is a great amount of research describing how ICT is being used effectively and efficiently in schools. DFES (2003) set out the aims for effective use of ICT in teaching and learning as broadening prospects with more opportunities for creative expression, flexibility to study when, where and how best suits the need and preferences of individual, increased motivation through learning that stimulates interest, wider accessibility to participation and learning, reasonable choices about when, when not and how to use new technology to enhance, improve and sustain learning and teaching. It further suggests that ICT can make essential contribution to teaching and learning across all subjects and ages. Thus it can engage, and inspire children and young people and meet their individual needs and preference. Cox (1999) also suggested some benefits of using ICT in lessons:

- Increased commitment to learning tasks/jobs
- Improved enjoyment and interest in learning the subject
- Increased in self –directed independent learning
- Enhanced self-esteem leading to expectations of achieving goals.

Becker (2001) documented a study of over 4000 teachers in the USA and suggested the following objectives of using ICT in lessons

- Getting ideas and information
- Expressing self in writing
- Understanding subject skills just taught
- Learning computer skills and
- Analyzing information

Students can derive from the effective and efficient use of ICTs in the teaching and learning process such as increased motivation to stay on-task, behave well and produce higher quality output, learn more individual and at their own pace, do things they cannot do using traditional methods and resources and finally, a combination of several subjects into project-based activities.

6.1 Identified barriers to ICT adoption in education at other places

Information Technology integration into teaching and learning is the application of technology to help, enhance, and speed-up student knowledge (Omwenga et. al, 2004). This means more than simply teaching learners how to use computers. The technology is rather a means for improvement in education and not an end in itself. Muriithi (2005) has argued that in Kenya, like most developing countries ICT is still at the awareness level limiting its application to basic computer literacy and office clerical tasks. A study carried out by Organization for Economic Cooperation Development (OECD) in 2009 and cited in Rodden (2010) confirmed that there are a number of factors that inhibit the use of ICT in education. These barriers included inconsistent number of computers to students, a deficit in maintenance and technical assistance and finally, a lack of computer skills and/or knowledge among teachers (OECD, 2009). Jenson et. al (2002) classified these barriers as limited equipment, inadequate skills, minimal support, time constraints and lack of interest or knowledge by teachers. In a related research report carried out by British Educational Communications and Technology Agency (BECTA) in 2004, some important barriers were identified. These were lack of confidence, accessibility, lack of time, fear of change, poor appreciation of the benefits of ICT and age. Ertmer (1999) concurs with Schoepp (2005), asserting that if teachers are sensitized or informed of such barriers, they can introduce strategies to overcome them.

According to Iddrisu (2009) although important lessons may be learned from best practices around the world, there is no one formula for determining the finest level of ICT integration into the educational system. Significant challenges that policy makers, education administrators and other stakeholders need to consider include educational policy planning; infrastructure; language and content; capacity building; and financing. In fact it is a tall list of research on the challenges Ertmer (1999); Balanskat et. al (2006); Pelgrum (2001); BECTA (2004). As stated by Iddrisu (2009), there is no any single formula but strategies with local content will be the best approach to integrating ICT into teaching and learning. In this research, we have leveraged importance on teacher related barriers that bother much of this research's objectives.

6.1.1 Teacher related barriers

The teacher is the principal stakeholder in the teaching and learning process and crucial in determining ICT use in the classroom

(Baylor and Ritchies, 2002). Gressard and Loyd (1985) put it to all that teachers' attitude towards ICT is a key factor which determines successful integration, while Jegede (2008) identifies the teacher as a key instigator in fostering ICT integration in teaching and learning. Teacher related barriers are summarized as:

i. Lack of knowledge or competence. A teacher's lack of knowledge serves as a considerable challenge to the use of computers in teaching methods and practices. Tezci (2009) as cited in Rodden (2010) posits that if teachers have a high level of ICT knowledge, then there will be a higher level of ICT use in education. These barriers according to some researchers vary from country to country. Pelgrum (2001) found that lack of knowledge/competence in technology, among teachers in developing nations, is the primary obstacle to the uptake of ICT in education.

ii. Lack of confidence
Several studies conducted reveal that the lack of confidence prevents teachers from using ICTs. A BECTA report in 2004 adumbrates that many teachers who are untrained in ICT are not prepared to use them in the classroom or in front of students who might probably know more than them. According to Jegede et. al (2008) as teachers become more grateful of the use of ICTs as a pedagogical aid, attitudes and interest mostly become positive. Causes of lack of confidence include fear (Beggs, 2000) whilst Balanskat et. al (2006) attribute it to limited ICT knowledge of the teacher.

iii. Fear for change
Computer fear is a key barrier, limiting or preventing the use of ICT by teachers according to a BECTA(2004) report.

iv. Lack of training
Most researchers have identified this barrier frequently cited Pelgrum (2001); Rodden (2010); Bingimlas (2009); BECTA (2004); Trotter (1999); Gomes (2005). Osborne and Hennessy (2003) found that teacher training is essential if they are to integrate new tools and approaches in education.

v. Age
Kumar et. al (2008) maintain that age is an important factor to the use of ICT in teaching and learning process. Lee (1997) points out that many older teachers did not have any computer education when in school and as a result are in need of training to allow them to make use of computers in their work. Cavas et. al (2009) conclude that there is a connection between teacher's age and their computer attitudes. Another study by Korte and Husing (2007) concludes that younger teachers appear to be less worried about using ICTs in learning.

6.2 Government policy on ICT adoption in Ghanaian schools.

The government of Ghana is poised for transforming the agro-based economy of the country into knowledge-based economy (Dzidonu et.al, 2003). The need for ICT-based training in the tertiary institutions, especially the polytechnics that are mandated to produce practical skills-oriented graduates, has been dominating government talk shows. The government acknowledges that the integration of ICT into the Education system will result in the creation of new opportunities for learners and teachers to engage in new ways of knowledge acquisition hence the rationale behind its ICT policy statement which is an epitomized version of the ultimate goal to transform the educational system. The policy document provides a clear policy direction for what needs to be done and how it is intended to be done. Unfortunately, programmes of implementation of the outlined policy actions have so far been crawling.

7. RESULTS AND DISCUSSIONS

The demographic characteristics and background information of the respondents were first looked at in the obtained data. This was followed with the discussions of the findings in connection with the research objectives.

7.1 Demographic characteristics and background information

The results obtained indicated that 78.8% of the respondents were males whilst the remaining 21.2% were females (Table 2). It was quite interesting to note that majority of the respondents aged between 46 and 60 years. This further implies that the polytechnic should be replacing majority of its teachers in the next 14 years (figure 1).

Table 2: Gender of respondents

Gender	Frequency	Percent
Male	63	78.8
Female	17	21.2
Total	**80**	**100.0**

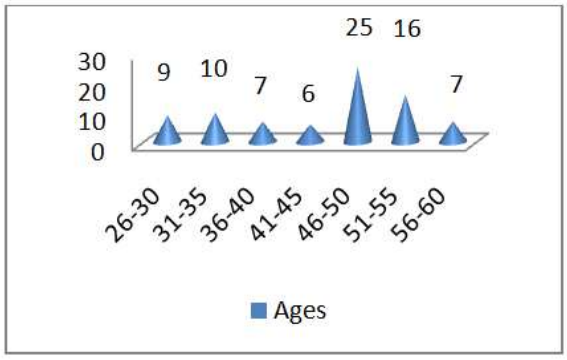


Figure 1 Ages of respondents

A scrutiny of the collected data summarized in table 3 revealed that the polytechnic has almost got the full compliments of lecturers. A total of 62 of the lecturers representing 77.5% are master degree holders. A few lecturers are undergrads whilst three (3) were terminal PhD holders.

Table 3: Highest Level of Education Attained

Certificate	Frequency	Percent
HND	5	6.3
1st Degree	10	12.5
Masters	62	77.5
Doctorate	3	3.8
Total	**80**	**100.0**

The study discovered that more than 50% of the polytechnic lecturers have a minimum of six (6) years experience in the polytechnic. The old adage that "experience is the best teacher" could therefore be put to bare or serve better in the polytechnic in terms of imparting knowledge to learners.

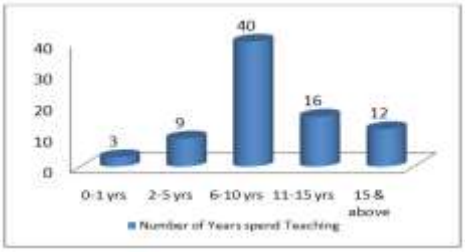


Figure 2 Teaching Experience

7.2 What are the factors inhibiting polytechnic teachers' use of ICT in teaching and learning?

The main objective of this work was to find the general answer to this question. A multi-response analyses of the respondents answers is summarized in table 4

Table 4 Factors preventing teachers from using ICTs in their lessons in classrooms

Cases	Responses		
	Frequency	Percent	Rank
Inadequate LCD projectors in classroom	46	12.2%	2
Lack of knowledge about computers	59	15.7%	1
Lack of confidence	32	8.5%	5
Fear for change	24	6.4%	8
Lack of training	43	11.4%	3
My age	37	9.8%	4
Little previous knowledge	23	6.1%	9
Not sure how useful computer are	20	5.3%	11
Computers are not accessible	26	6.9%	7
Management doesn't care if I use computer or not	28	7.4%	6
Computer equipment is unreliable	17	4.5%	12
No support if something goes wrong with computer	21	5.6%	10
Total	**376**	**100.0%**	

In a multi-response questioning, the cases are not mutually exclusive. In this regard, a total of 376 cases were obtained as seen in the above table. The three top responses came from Lack of knowledge about computers with 15.7%; Inadequate LCD projectors in classrooms representing 12.2%; and Lack of training being 11.4%. A total of 37 (9.8%) lecturers think they are too old to embrace the change in teaching methodology. The findings also point to a fact that the "old age" are the likely respondents who lack confidence (8.5%) in using computers to teach. Further exploration of the responses shows that management is interested in the adoption of ICT in teaching and learning since only 7.4% think that Management does not care if they use computers. Another interesting finding is that the polytechnic has good technical support team. This is appreciated from the 5.6% who think they would not get technical support in case of computers malfunctions. Notions of fear for change in the teaching were appreciated though a minority (6.4%). Interestingly, a few respondents consider the computers to be unreliable (4.5%). Indeed, some inactive users of computers usually have this reservation.

In fact one can conclude as a deduction from the above expositions that the three (3) major barriers preventing the use of ICTs by teachers in Tamale Polytechnic are

i) Lack of knowledge about computers
ii) Inadequate LCD projectors in classrooms
iii) Lack of training being

7.3 Availability of ICT tools or equipment in the school.

In trying to address this specific objective, respondents were asked whether they had sufficient computers, accessories and LCD projectors for use in their departments. As can be appreciated in table 5, only eight lecturers constituting 10% answered affirmatively. It was very possible that these responses might have come from the Computer Science department that has a number of computers for computer literacy lessons. The general opinion is that computers and accessories for lecturers' use are inadequate or do not exist.

Table 5: Existence of sufficient computers and accessories for lecturers' use.

Response	Frequency	Percent
Yes	8	10.0
No	72	90.0
Total	**80**	**100.0**

7.4 Adequate training in ICT to deliver lectures using computers

ICT capacity building exercises for lecturers in the polytechnic were found to be lagging behind. Only fifteen (15) out of the eighty (80) respondents believed they had received sufficient ICT training.

Even though the institution's management is interested in adopting ICT for teaching and learning, very little has been done to strategically achieve this objective (figure 3).

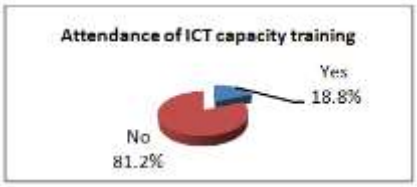


Figure 3 Attendance of ICT training

7.5 The attitudes of teachers in the use of computers

Attitude can be seen as a positive or negative sentiment according to Ajzen and Fishbein (1980). Smith et. al (2000) maintains that computer attitude evaluation usually encompasses statements that examine users' interaction with computer hardware, software, other persons relating to computers and activities that involve computer use. It was discovered that only eleven respondents (approximately 14 %) indicated that they use computers to prepare lecture notes and recording examination results. An interesting finding was that the remaining sixty nine (representing 86%) who have never interacted with computers were willing to use them if they obtained the requisite skills.

8. SUMMARY OF FINDINGS

The study came up with the following findings:
The major factors that inhibit teachers from using ICTs in Tamale Polytechnic are ranked below:

i. Lack of knowledge about computers
ii. Inadequate LCD projectors in the classroom
iii. Lack of ICT capacity building to lecturers
iv. Age of lecturers
v. Lack of confidence

The least factors that were identified are:

vi. Lack of Management pressure on lecturers to use computers for teaching and learning
vii. Computers are not accessible
viii. Fear for change
ix. Lack of previous ICT knowledge
x. Insufficient technician support if something goes wrong with the computers
xi. Not sure how useful computers are
xii. The myth that computers are unreliable.

A number of by-findings were made:

i. Majority of the Tamale Polytechnic lecturers are closed to retirement.
ii. Majority of the lecturers are second degree holders whilst a few with terminal PhDs.
iii. Interviews with stakeholders indicated that the polytechnic has just established an ICT Services Directorate and has developed an ICT strategic plan for the period of 1st January 2016 to 31st December 2019.
iv. The study revealed that even though government ICT policy on education provides a clear policy direction for what needs to be done, implementation has been very poor.

9. RECOMMENDATIONS AND FUTURE ENHANCEMENT

The research findings suggest the following recommendations:
The management of the polytechnic should intensify ICT adoption campaign through sensitization and in-house capacity building workshops. This can be done easily using the lecturers from the Computer Science Department in collaboration with the ICT Directorate. Lack of the necessary competencies breeds lack of confidence and fear for change. In fact age affecting ICT competence is a wrong notion or a myth. Adequate trainings can resolve all these hindrances, making the lecturers abreast with the modern pedagogy of imparting knowledge and skills.
Committed funds from the polytechnic's IGF (Internally Generated Funds) and efforts should be exerted on acquiring laptops, LCD projectors and other equipment for audio visual teaching and learning. For example, acquisition of fifteen (15) laptops and fifteen (15) LCD projectors annually will allow the polytechnic to cover all its fifty eight (58) lecture halls with PowerPoint-enabled teaching aids.
Computer literacy should be made to score high marks in new lecturers' recruitment.
Polytechnic administrators should pay more attention to ICT investments in their business to derive the most. The way forward is IT and should be seen as such.
Governments of developing countries and for that matter Ghana, should now take a pragmatic approach towards realizing their ICT dreams. In most of these countries, public funds have been spent to develop very nice ICT policies and plans but implementations fail. A research to ascertain why the implementations of such plans in developing countries fail is highly recommended.

10. REFERENCES

[1] Balanskat, A., R. Blamire and S Kefala. "A review of studies of ICT impact on schools in Europe." European Schoolnet: European Communities (2006).
[2] Baylor, A. L. and D. Ritchie. "What factors facilitate teacher skill, teacher morale, and perceived student learning in technology-using classroom?" Computers & Education (2002): 39(4), 395-414.
[3] Becker, Henry Jay. "How Are Teachers Using Computers in Instruction?" 2001 Meetings of the American Educational Research Association. Seattle: American Educational Research Association, 2001.
[4] BECTA. "A Review of the Research Literature on Barriers to the Uptake of ICT by Teachers." 2004. 19 Dec 2009 <http://partners.becta.org.uk/page_documents/research/barriers.pdf>.
[5] Beggs, T.,A. "Influences and barriers to the adoption of instructional technology." 2009. 19 Dec 2009 <http://www.mtsu.edu/~itconf/proceed00/beggs/beggs.htm>.
[6] Bigum, Chris. "Teachers and computers: in control or being controlled?" Australian Journal of Education (1997): 41 (3): 247-261.
[7] Bingimlas, K. "Barriers to the Successful Integration of ICT in Teaching and Learning Environments: A Review of the Literature'." Eurasia Journal of Mathematics, Science and Technology Education (2009): 5(3), 235-245.

[8] Busha, Charles H. and Stephen P. Harter. Research Methods in Librarianship: Techniques and Interpretation. Michigan: Academic Press, 1980.
[9] Cavas, B., et al. "'A Study on Science Teachers' Attitudes Toward Information and Communication Technologies in Education'." The Turkish Online Journal of Educational Technology (2009): 8(2), 20-32.
[10] CISSSE, Lamine. "ECOWAS and the Daily Events: The Present Realities of the Integration Process within the Sub-Region ." GJA and FES (2006): 47.
[11] Cox, M.J. "What factors support or prevent teachers from using ICT in their classrooms?" British Educational Research Association Annual Conference. Brighton, 2000.
[12] DFES. The big pICTure: the impact of ICT on Attainment, Motivation and Learning. Nottingham: DfES, 2003.
[13] Drent, Marjolein and Martina Meelissen. "'Which Factors Obstruct or Stimulate Teacher Educators to Use ICT Innovately?'." Computers and Education (2008): 51(1), 187-199.
[14] Dzidonu, Clement K. and et.al. The Ghana ICT for Accelerated Development Policy. Accra: Ministry of Communication, 2003.
[15] Ertmer, P. "'Addressing first-and second-order barriers to change Strategies for technology integration'." Educational Technology Research and Development (1999): 47(4), 47-61.
[16] Gomes, C. "Integration of ICT in science teaching. A study performed in Azores, Portugal." Recent Research Developments in Learning Technologies (2005).
[17] Gregoire, R., R. Bracewell and T. Laferriere. "The Contribution of New Technologies to learning and Technology in Elementary and Secondary School." (1996).
[18] Gressard, C., P. and B., H. Loyd. "'Age and staff development experience with computers as factors affecting teacher attitudes towards computers'." School Science and Mathematics (1985): 85(3), 203-209.
[19] Hawkridge, D. "(1990). Computers in Third World Schools. The example of China." British journal of educational technology (1990): 21 (I): 4-20.
[20] Hepp, P., et al. Technology in schools: Education, ICT and the knowledge society. : World. Washington, DC: The World Bank, 2004.
[21] Iddrisu, S.,A. Predictive validity of Senior High School (SHS) aggregate of students' grade-point average (gpa). Cape Coast, Ghana: Unpublished doctoral dissertation submitted to the Department of Educational Foundations, Faculty of Education, University of Cape Coast, 2009.
[22] Jegede, P., O. Dibu-Ojerinde and M. Llori. "'Relationships Between ICT Competence and Attitude Among Nigerian Tertiary Institution Lecturers'." Educational Research and Review (2007): 2(7), 172-175.
[23] Jenson, J., B. Lewis and R. Smith. "'No one way: Working models for teachers' professional development'." Journal of Technology and Teacher Education (2002): 10(4), 481-496.
[24] Korte, W., B. and T. Husing. "'Benchmarking Access and Use of ICT in European Schools 2006:Results from Head Teacher and A Classroom Teacher Surveys in 27 European Countries'." eLearning Papers (2007): 2(1), 1-6.
[25] Kozma, Robert B. "Will Media influence learning? Reframing the debate." Educational TechnologyRresearch and Development (1994): 42 (2): 7-19.
[26] Kozma, Robert, B. "Comparative Analysis of Policies for ICT in Education." Internatonal Handbook on Information Technology in Education (2008).
[27] Kumar, N., R.,C. Rose and J.,L. D'SIiva. "Teachers' Readiness to Use Technology in the Classroom: An Empirical Study." European Journal of Scientific Research (2008): 21(4), pp. 603-616. .
[28] Lee, D. "'Factors influencing the success of computer skills learning among in-service teachers'." British Journal of Educational Technology (1997): 28(2), 139-141.
[29] Lowther, D. L., et al. "Does technology integration work when key barriers are removed?" Educational Media International (2008): 195-213.
[30] Matthew, K. "A Comparison of the Influence of Interactive CD-ROM Storybooks and Traditional Print Storybooks on Reading Comprehension." Journal of Research on Computing in Education (1997): 263–275.
[31] McGrail, E. "Teachers, Technology, and Change: English Teachers' Perspectives." Journal of Technology and Teacher Education, Norfolk, VA 13.1 (2005): 13 (1), pp. 5-24.
[32] MIE. Participatory teaching and learning, a guide to methods and techniques. Domasi, Malawi: Malawi Institute of Education, 2004.
[33] Muriithi, P. A framework for integrating ICT in the teaching and learning process in secondary schools in Kenya. Nairobi: Unpublished MSc. Thesis submitted at the University of Nairobi, School of computing andInformatics, 2005.
[34] OECD. "Education At a Glance: 2009 OECD Indicators." Organization for Economic Cooperation Development Publishing (2009).
[35] Omwenga, Elijah, T. Waema and P. Wagacha. "A model for introducing and implementing e-learning for delivery of educational content within the African context." African Journal of Sciences and Technology (2004): 5 (1), 35-48.
[36] Osborne, Jonathon and Sara Hennessy. "Literature Review in Science Education and the Role of ICT: Promise, Problems and Future Directions." 2003.
[37] Pelgrum, W. and N. Law. ICT in Education around the World: Trends, Problems and Prospects. Paris: UNESCO, 2003.
[38] Pelgrum, W., J. "'Obstacles to the Integration ICT in Education: Results from a Worldwide Educational Assessment'." Computers and Education (2001): 37(2), 163-178.
[39] Rodden, N. An investigation into the barriers associated with ICT use in the Youthreach classroom: A case study of a centre for education in the North West. Limerick: Unpublished Master"s thesis,University of Limerick, 2010.
[40] Schoepp, K. "'Barriers to Technology Integration in a Technology-Rich Environment Learning and Teaching

in Higher Education: Gulf Perspectives." Learning and Teaching in Higher Education: Gulf Perspectives (2005): 2(1), 1-24.

[41] Sproull, Natalie, L. Handbook of research methods: A Guide for Practitioners and Students in the Social Sciences. Second Edition. Metuchen, N.J., & London: The Scarecrow Press, Inc., 1995.

[42] Tezci, E. "'Teachers' effect on ICT use in education: the Turkey sample', ." Procedia Social and Behavioral Sciences (2009): (1)1, 1285-1294.

[43] Trotter, A. "Preparing teachers for the digital age." 2009. 5 Sept 2009 <http://oak.cats.ohiou.edu/~waltje/classes/media2008/preparing.pdf>.

[44] Weert, T. V. and A. Tatnall. "Information and Communication Technologies and Real-Life Learning: New Education for the New Knowledge Society." Springer (2005).

[45] Yamane, Taro. "Statistics: an introductory analysis.". New York: Harper & Row, 1973.

International Journal of Computer Applications Technology and Research
Volume 5–Issue 9, 603-611, 2016, ISSN:-2319–8656

Segmentation and Visualization of Human Coronary Artery Trees from CTA Datasets

Qian Huang

Computer Science Department
Wright State University, United States

Abstract: The volume information extracted from computed tomography angiogram is very useful for cardiologists to diagnose various diseases. An approach is presented to segment human coronary artery trees from the volumetric datasets. The coronary arteries' surfaces are recovered by triangle mesh with the boundary points extracted from the coronary artery voxels segmented. The positions where the calcified plaques occur are identified by mapping the intensities of boundary points of the coronary artery trees on the triangle meshed surfaces. If different values of the computed maximum principle curvatures of boundary points surrounding the lumen cross section are mapped on the triangle meshed surfaces of the segmented coronary artery trees, the cross section structure of the coronary artery lumen segment is noncircular cross section structure.

Keywords: **Segmentation, visualization, coronary arteries, calcified plaque, maximum principle curvature.**

1. INTRODUCTION

Computed tomography angiogram (CTA) [1] plays a significant role as a clinical tool in the diagnosis of the coronary artery diseases. It provides three-dimensional information [2], allows for a better understanding of cardiac three-dimensional anatomy [4], [5], medical diagnosis [6], and ongoing investigations of acute and chronic coronary heart diseases [7]. The noninvasive [9] manner empowered with the visualization of vessels attracts cardiologists to apply it in clinical environments. Segmenting and visualizing coronary artery trees and deriving quantitative data [18] from CTA image datasets for inspecting the coronary lesions is what many algorithms' goal to focus on.

Several algorithms have been developed to segment and visualize vessels in three dimensions [3], for instance, level set method [10], active contour algorithm [14], vesselness measurement [11], [12], expectation maximization estimation algorithm [15], moment-based shape analysis for voxel clusters [16], shape model based algorithms such as tubular model in three dimensions [17], the algorithm of combining graph-cuts and robust kernel regression to segment coronary lumens [18], [19]. The standard marching cube algorithm [21] generates a high resolution isosurface with an isovalue of image intensity to represent the object's surface. The vertices of the set of triangles build the boundary point cloud of the object. A gradient based algorithm is used to extract vessel boundaries, the vessel boundaries are represented as an unstructured point cloud in three-dimensional image space. The locations of the vessel boundaries are defined among a set of voxels with the maximum gradient magnitude along the gradient direction. In this work, this algorithm will be applied to produce a boundary point cloud representing the coronary artery boundaries of a human heart.

The existences of the calcified and soft plaques are harmful to the health of a human body [22-24]. Calcified plaques are with high intensities [15]. Assuming voxels' intensities of the segmented coronary arteries are a Gaussian distribution brings out calcified plaques' being identified. The computed mean value plus three standard deviations of the intensities is the threshold value to recognize the calcified plaques [25]. Stenoses or soft plaques are studied to be detected with profiles of artery lumen sectional area or vessel radiuses along the vessels' centerlines [15], [27], [30]. The detections of the artery lumen, calcified and soft plaques or stenoses are clinical useful [31-33]. A coronary artery segmentation method is presented in this work to segment the voxels representing the coronary artery trees from CTA datasets. A triangle mesh method is used to recover the coronary artery surface from those detected vessels' boundary points. The positions of calcified plaques can be identified from the recovered triangle surfaces. The computed maximum principle curvatures of the coronary boundary points are mapped on the triangle surfaces.

2. SEGMENT THE CORONARY ARTERIES WITH MULTIPLE ISOVALUES

2.1 Downsample the CTA Datasets

Similarly as [8], the sampling theorem in three dimensions is applied in this work. The discrete image dataset is obtained by sampling a continuous function $f_c(x, y, z)$ on a lattice with interval (X, Y, Z). For different sampling rate of a continuous function, the lattice interval (X, Y, Z) takes different values. Different lattice interval value represents different voxel size. Then for the same continuous function, the discrete image dataset, when sampled at lower sampling rate, has larger lattice interval, therefore larger voxel size. For the same spatial dimensions, if the discrete image dataset is sampled at high sampling rate, the voxel size is smaller, and the number of the resulted voxels is greater than those sampled at low sampling rate.

According to marching cube algorithm, to compute the vessel boundary points from the original size of the CTA dataset, each voxel has to be computed. The extracted boundary point cloud includes the boundary points of arteries, chambers and other tissue structures. In a discrete image dataset, the vessels and other tissue structures are expressed as connected voxels. For the original image dataset, the number of voxels representing the vessels and other tissue structures is greater than those of image dataset with low sampling rate. Then the resulted boundary points from the original size of CTA dataset are more than those from the downsampled CTA dataset.

Consequently, the amount of memory space involved for computing a downsampled CTA dataset is decreased. To downsample the image dataset, the image dataset is convolved with a sinc function and the original continuous function is

reconstructed, so that it can be resampled. The sinc function has infinite support, therefore, the windowed sinc function has to be used, which truncates the sinc function by multiplying it with a window function. After downsampling the volumetric image dataset, the size of the dataset is changed to 1/3 of the original image size in both rows and columns, respectively, and 1/1.6 of the original image dataset size in the number of slices.

2.2 Detect Coronary Artery Regions

Computation of all voxels in each image slice will produce many boundary points including other tissue structures besides the boundaries of the vessels, and the involved memory is of large amount. It is better to segment the regions containing coronary arteries for the extraction of their boundary points. Mathematical morphological operators can cluster a set of connected pixels with a structure element. In this work, such operators are applied to detect the coronary artery regions.

Dilation morphological operation is to move a structure element inside the region of interest, and outputs the locus of the pixels covered by the structure element when its center moves inside the region of interest. The intensities of the artery regions are higher compared with those of the heart muscles. For each image slice, its average intensity is computed. If the pixels' intensities in each image slice are higher than the average intensity plus or minus a small offset defined by the user for segmenting different coronary artery trees, the center of a 3x3 square structure element is placed on them. All the pixels covered by this square are defined as the foreground region. Erode this foreground region with an 8x8 square when the downsampled dataset is being processed and with a 24x24 square when an original dataset is being operated. Small regions of foreground voxels can be removed via this operation. The resulted erosion set is dilated with a 4x4 square (with a downsampled dataset) or 12x12 square (with an original dataset) again. The difference of the foreground region and the resulted set is the region where coronary arteries exist. With this region, the produced boundaries belong to other tissue structures and the involved memory can be reduced. An image slice as an instance of the coronary artery region detected is shown in Figure. 1 (a).

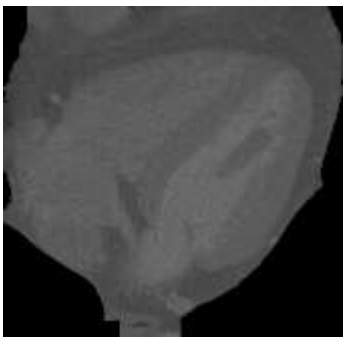
2.3 Analyze Image Histograms of the CTA Datasets with the Original Size

The image displays an object when its average intensity is different from its adjacent areas. The histogram provides an image's gray-level distribution. Multimodal histograms can occur when the image contains multiple objects of different average brightness. The CTA datasets' intensity histograms are shown in Figure. 2. There are three peaks in both of the histograms. The first one represents the background objects and its intensity is the lowest. The second one indicates the blood's intensity and its intensity is between the lowest and the highest. The third one is recognized as the objects with the highest intensities, for instance, the bones or calcified plaques, etc.

2.4 Segment with Multiple Isovalues

The standard marching cube algorithm is applied to extract an object's boundary point cloud. It is to find a boundary point linearly interpolated along the edge between two adjacent voxels in x, y, or z direction. Thus the object's boundary points generated by the standard marching cube algorithm vary with the intensity threshold value defined by the user. Since the varying contrast agent among the coronary arteries, multiple threshold values, similarly as in [26], are used in this work to extract coronary artery boundaries in the marching cube algorithm. The tested CTA datasets' intensities are ranged from 0 to 4095. These multiple threshold values are selected inside an

(a)



(b)

Figure. 1 (a) an image slice with the coronary artery region detected, (b) an original image slice without detection operation

intensity region around the second peak value in the CTA dataset's intensity histogram. The intensity threshold values are listed in Table I.

TABLE I THE MULTIPLE INTENSITY THRESHOLD VALUES

Dataset	Intensity values			
1	1022	1102	1145	1187
2	1343	1445	1500	1562.5

Since multiple thresholds are used, multiple boundary layers are perhaps produced for the vessels. To reduce multiple boundary layers to one layer, the point with maximum gradient magnitude is kept among the neighboring boundary points near the gradient line of each boundary point.

2.5 Determine the Coronary Arteries' Centerlines

The computed boundary point cloud is tetrahedralized, centerpoints can be calculated by applying a topological analysis of the vector field based on the gradient field within every tetrahedron. The computed centerpoints can be connected via a line segment when their nearest neighboring centerpoints are found. The centerpoint can be connected with another centerpoint that is also closest to it and is in the reverse direction. Thus this centerpoint's neighbors are determined.

To differentiate the coronary artery centerpoints from the computed centerpoint cloud, one experienced medical technician can pick one seed centerpoint among the centerpoint cloud. This centerpoint is recognized as a coronary artery centerpoint; the neighboring centerpoints, which it is connected to, are recursively marked as coronary artery centerpoints. Several seed

centerpoints with a downsampled dataset are selected to find the other portions of a coronary artery tree's centerlines. The number of seed centerpoints with an original dataset is more than a downsampled dataset.

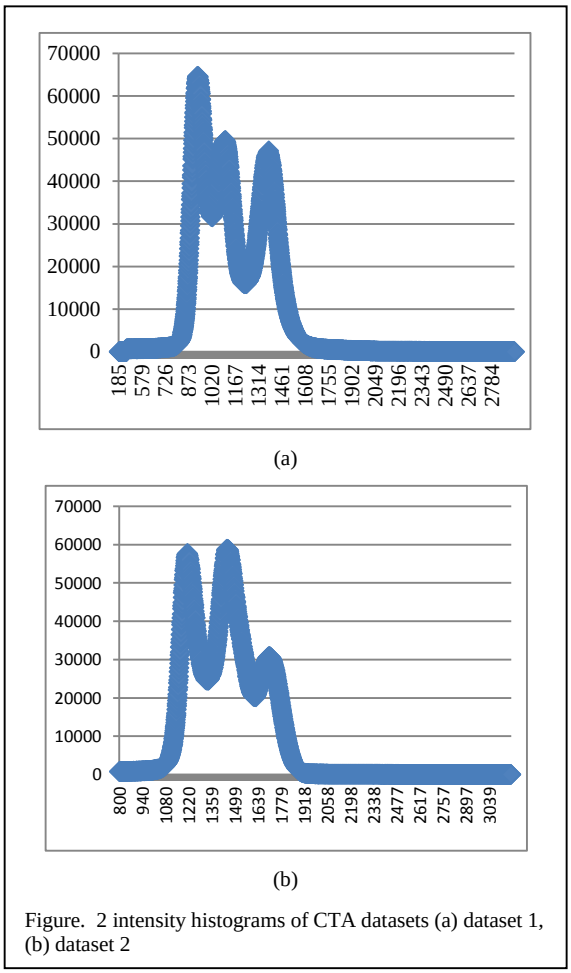


Figure. 2 intensity histograms of CTA datasets (a) dataset 1, (b) dataset 2

2.6 Find the Coronary Artery Boundaries

For each coronary artery centerpoint, the boundary points are recognized as boundary points of the coronary arteries when they are enclosed in a ball with the radius of current centerpoint. These boundary points' normals must point inward of the vessel that encloses the current centerpoint. The tetrahedra attached by each of these boundary points are defined as the vessel tetrahedra except those, of which, there are any vertices' normals pointing outward of the vessel. The vertices of vessel tetrahedra are identified as the boundary points of the coronary arteries.

2.7 Segment the CTA Dataset with the Original Size

The coronary artery boundary voxels of the original CTA dataset can be located from those coronary artery boundary points from the downsampled CTA dataset and are dilated with a 3x3x3 cube when a right artery tree is segmented and a 5x5x5 cube when a left artery tree is segmented. If the voxels belong to the detected coronary artery region, they are again processed with the same procedures to the downsampled datasets, for instance, segmenting with multiple isovalues, tetrahedralizing the boundary points, determining the coronary artery centerlines and finding the coronary artery boundaries. To extract the coronary artery centerlines, the centerlines represent small vessel branches are removed manually. Thus the vessel tetrahera are defined and the coronary artery boundary points are found.

A scan-conversion algorithm [20] for lines computes the coordinates of the voxels that line on or near an ideal, infinitely thin straight line in three dimensions. In principle, the sequence of voxels is required to lie as close to the ideal line as possible and to be as straight as possible. The line segment can be drawn with a sequence of voxels described above with given start point and end point of this line segment. Scan converting polygons as area-defining primitives could be done a line segment at a time. Similarly, scan converting a tetrahedron as a volume-defining primitive can be completed a triangle at a time. Then the voxels representing a tetrahedron can be identified. And the coronary artery voxels can be obtained from all the vessel tetrahedra. The coronary artery boundary points are triangulated with a surface mesh. And the coronary artery voxels also include the voxels by the scan converting each triangle.

3. VISUALIZE THE CORONARY ARTERY TREES

The intensities of voxels that are not coronary artery voxels are set to zero. Those voxel sets of coronary arteries are smoothed with recursive Gaussian filter, and then are processed with the gradient based extraction algorithm. The triangle mesh is applied to the resulted boundary points to recover the surfaces of the segmented coronary artery trees.

3.1 Rendering the Surfaces of the Coronary Artery Trees

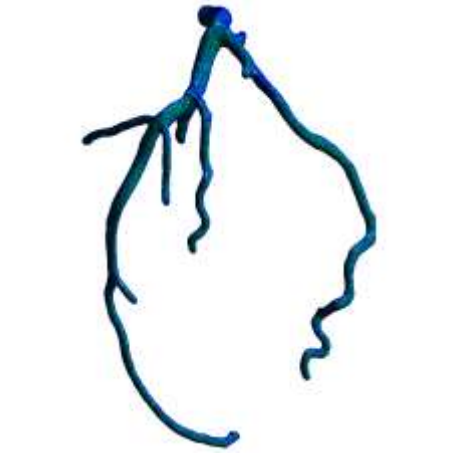
Two main visualization techniques for three-dimensional medical images are surface rendering and direct volume rendering (DVR) [13]. Rendering an object' surface with triangle mesh belongs to the surface rendering visualization category. In DVR, a projected two-dimensional image represents the entire three-dimensional dataset. The projected two-dimensional image does not allow DVR to capture all the three-dimensional surface information completely. Figure. 3 (a), (b), (c), (d) display the surfaces recovered by the triangle mesh on the segmented coronary artery trees. The number of boundary points, the triangles and hole triangles are listed in the Table II, 1 represents dataset 1, 2 indicates dataset 2. A hole triangle means at least one edge of the triangle connects only one triangle.

TABLE II. THE TRIANGLE MESH WITH THE ARTERY TREES' BOUNDARY POINTS

Artery tree	point number	triangle number	Hole triangle number
Left 1	32481	64961	14
Right 1	33694	67400	0
Left 2	44189	88376	12
Right 2	27008	54008	9

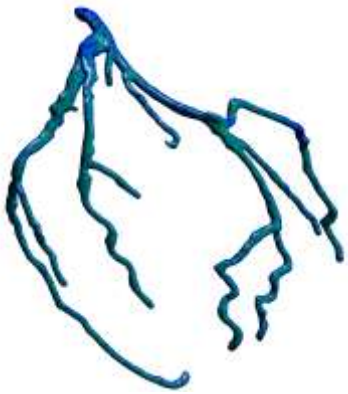
3.2 Identify Calcified Plaques on the Triangle Meshed Surfaces of the Coronary Artery Trees

For the varying contrast agent among the coronary arteries, intensities of the boundary points are assumed as with Gaussian distribution. The mean ($\mu_{boundary\ point}$) and standard deviation



(a)

(b)



(c)

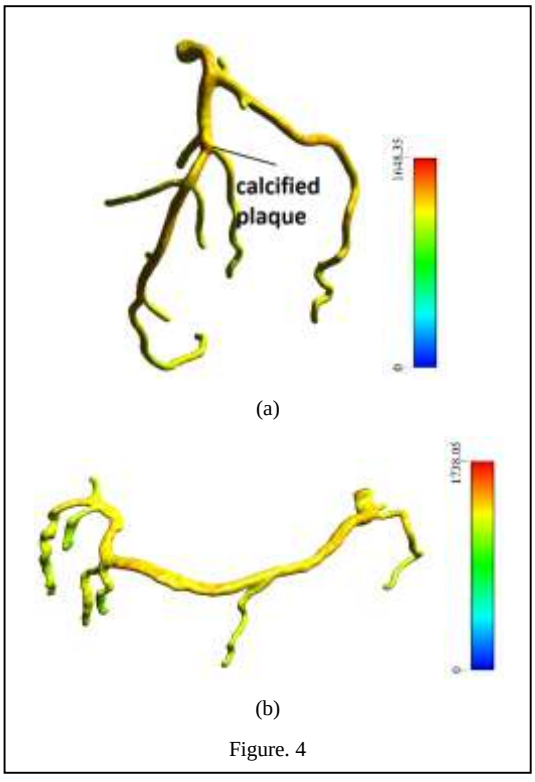
(d)

Figure. 3 triangle mesh on the artery trees (a) left 1, (b) right 1, (c) left 2, (d) right 2, 1 represents dataset 1, 2 indicates dataset 2.

($\sigma_{boundary\ point}$) of intensities of the boundary points are computed. Figure. 4 shows the intensities of the artery trees' boundary points mapped on the triangle mesh. The intensity values are displayed varying from low to high with color ramping map. The calcified plaques are having higher intensities compared with their neighboring intensities and are included in the vessel lumens by the multiple isovalue based algorithm. The color red represents those boundary points with intensities higher than $\mu_{boundary\ point} + 4\sigma_{boundary\ point}$. The calcified plaques can be perceived as a small block with the color red or light red if that light red's surrounding areas are colored green. Thus the calcified plaques with intensities lower than the threshold values can be identified. These calcified plaques are identified correctly as those are recognized from each image slice when the intensities of the boundary points are mapped on the triangle mesh.

Figure. 5 shows instances of the calcified plaques recognized from image slices circled with orange color. Figure. 5 (c) also displays one calcified area circled with blue color, which is not recognized with the boundary point intensity mapping, a false negative error. With tested coronary arteries, this calcified plaque detection method has zero false positive and only one false negative.

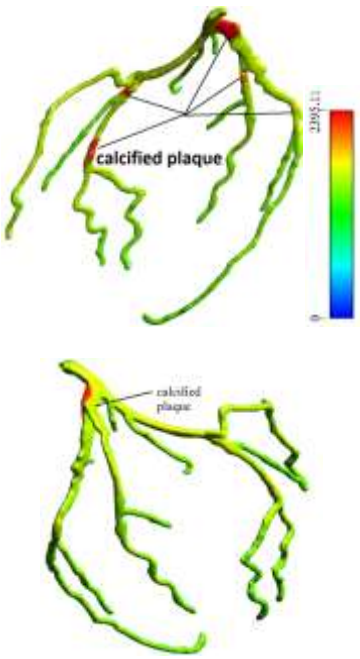
Figure. 6 visualizes the calcified plaques as sets of voxels colored orange, and they are recognized since they have higher average intensity than that of the blood of the coronary arteries. The calcified plaques are inside the vessel lumens by the multiple isovalue based algorithm and they are included in the segmented coronary artery voxels. Assume the blood intensities of the coronary arteries are with Gaussian distribution. The calcified plaques are recognized as those voxels which intensity values are greater than $\mu_{voxel} + 3\sigma_{voxel}$, μ_{voxel} is the mean intensity of the segmented coronary artery voxels, σ_{voxel} is the standard deviation. The calcified plaques with intensities less than the threshold value cannot be determined. In dataset 1 and 2, there



(a)

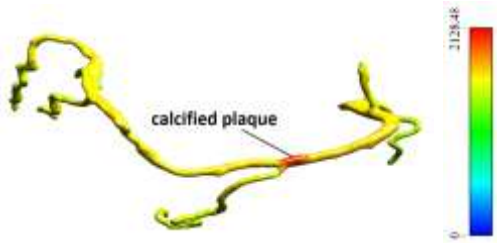
(b)

Figure. 4



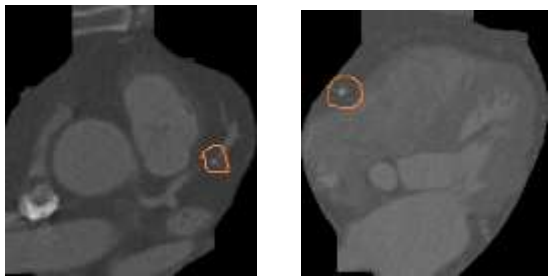
(c)

Figure. 4



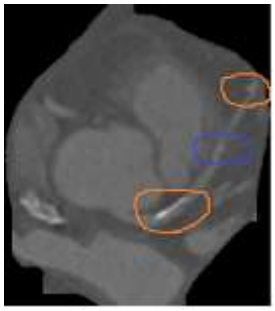
(d)

Figure. 4 intensities of the artery trees' boundary points mapped on the triangle meshed surfaces (a) left 1, (b) right 1, (c) left 2, (d) right 2, 1 represents dataset 1, 2 indicates dataset 2.



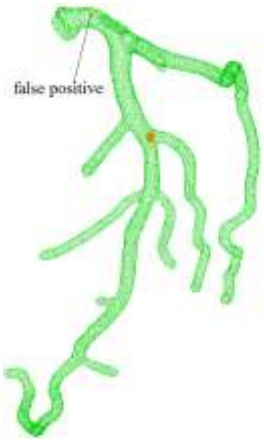
(a) (b)

Figure. 5



(c)

Figure. 5 calcified plaques recognized from image slices (a) left 1, (b) right 2, (c) left 2, 1 represents dataset 1, 2 indicates dataset 2.



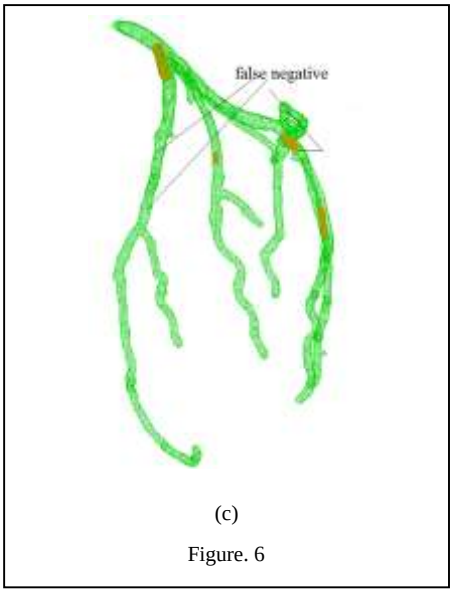
(a)

Figure. 6

(b)

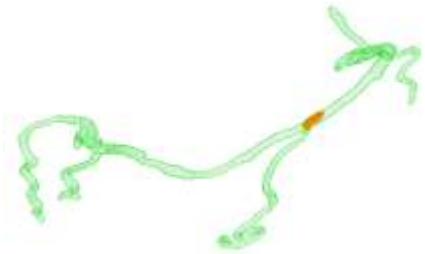
Figure. 6

is one false positive calcified area in the left coronary artery tree of dataset 1, and there are still three false negative areas in the left coronary artery tree of dataset 2.



(c)

Figure. 6



(d)

Figure. 6 calcified plaques visualized as orange voxels inside the coronary artery boundaries (a) left 1, (b) right 1, (c) left 2, (c) right 2, 1 represents dataset 1, 2 indicates dataset 2.

3.3 Mapping Maximum Principle Curvatures of Boundary Points on the Triangle Meshed Surfaces of the Coronary Artery Trees

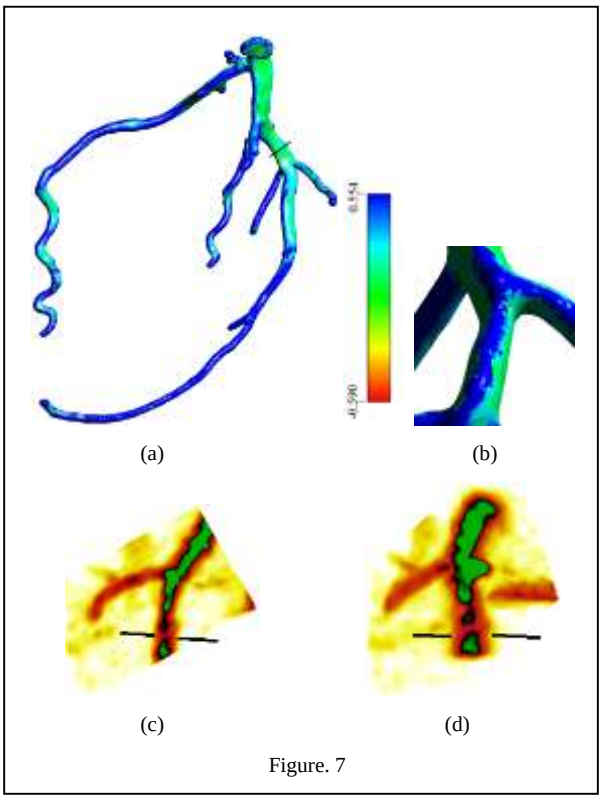
The maximum principal curvature of a torus is a constant which is equal to the reciprocal of the minor radius [29], and the maximum principal curvature of a circular cylinder is also a constant with a value of the reciprocal of the radius of circle base [29]. This work uses the hypotheses that if the real lumen cross section structure is deviated from the circular cross section structure, different values of the maximum principle curvatures of boundary points are surrounding the lumen cross section on the triangle meshed surfaces of coronary artery trees.

According to the curvature computation in [28], the maximum principal curvatures can be derived from the local Hessian matrix at each surface triangle's vertex. They are mapped on the recovered triangle meshed surfaces of the coronary artery trees. The mean value of those maximum principle curvatures of boundary points is computed. In Figure. 7 (a), (b), (e), (f), (i), (j), (m), (n), (q), (r), the color blue represents those values of maximum principle curvature that are greater than the mean value. Figure. 7 (b), (f), (j), (n) show the large views of the coronary artery segments located with line segments, different colors are displayed surrounding the lumen segments. The boundary points of these artery segments show light blue or green colors for low maximum principle curvature values and blue colors for high maximum principle curvature values. The positions of high (low) maximum principle curvatures are displayed at symmetric locations around the lumen segment's axis. These artery segments are investigated by the maximum intensity projection (MIP) technique. It is one of recommended image post processing formats for interpreting the coronary CTA dataset [34]. In this work, the MIP images are parallel projected. From the three-dimensional positions where the line segments are placed, the slab volumes containing the positioned artery segments can be found. Figure. 7 (c), (d), (g), (h), (k), (l), (o), (p) are the MIP images of the slab volumes containing the vessel segments that are located with line segments. Figure. 7 (c), (g), (k), (o) are the MIP images of one side view to project the slab volume onto the projection plane where the side of colors with high values of the maximum principle curvatures can be perceived. The MIP images of another view are shown in Figure. 7 (d), (h), (l), (p), they are created when the slab volume is projected onto the projection plane where the side of colors with low values of the maximum principle curvatures can be observed. The MIP images of these two side views illustrate different radii can be obtained from these two sides views at the same cross sections of the artery segments. The different radii displayed by these MIP images agree with the different colors mapped by maximum principle curvatures. (q) is another view of mapping maximum principle curvatures on the triangle meshed surface of left coronary artery of dataset 1. The artery segment positioned by the line segment is surrounding with blue color completely displayed in (r). The MIP images of the two side views of this artery segment in (s) and (t) display not obvious difference of radiuses at the same cross sections of the artery segment positioned by the line segment, which agree with the same color mapped by the maximum principle curvatures.

In this work, the computed maximum principle curvatures are inspected at positions without detected calcified plaques. The tested coronary artery segments located by the line segments show that if different values of the computed maximum principle curvatures of boundary points are surrounding the same lumen cross section on the triangle meshed surfaces of coronary artery trees, the lumen displays different radiuses surrounding the same cross section by MIP images, which further indicates this cross section structure is not a circular cross section structure.

4. CONCLUSIONS

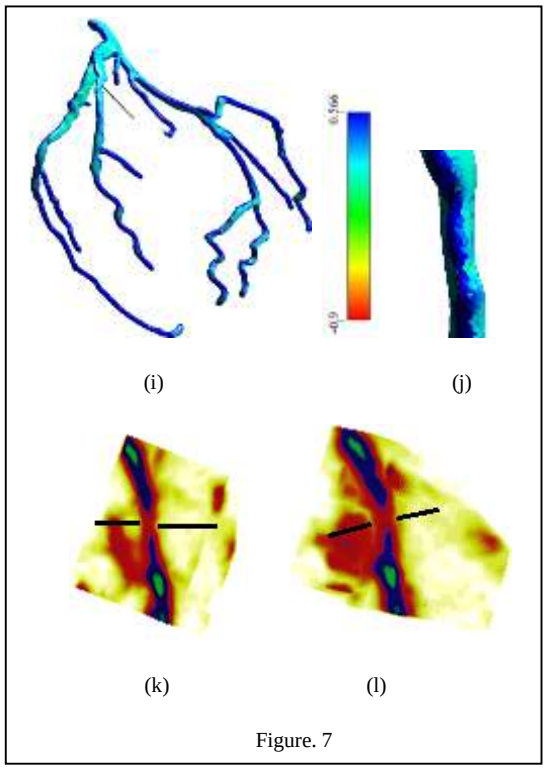
The method presented to segment coronary artery trees is capable to obtain the sets of voxels representing coronary arteries. The positions of the calcified plaques can be identified when the intensities of boundary points of the coronary artery trees are mapped on the triangle meshed surfaces. If different values of the computed maximum principle curvatures of boundary points surrounding the lumen cross section are mapped on the triangle meshed surfaces of the segmented coronary artery trees, the cross section structure of the coronary artery lumen segment is noncircular cross section structure.



(a) (b)

(c) (d)

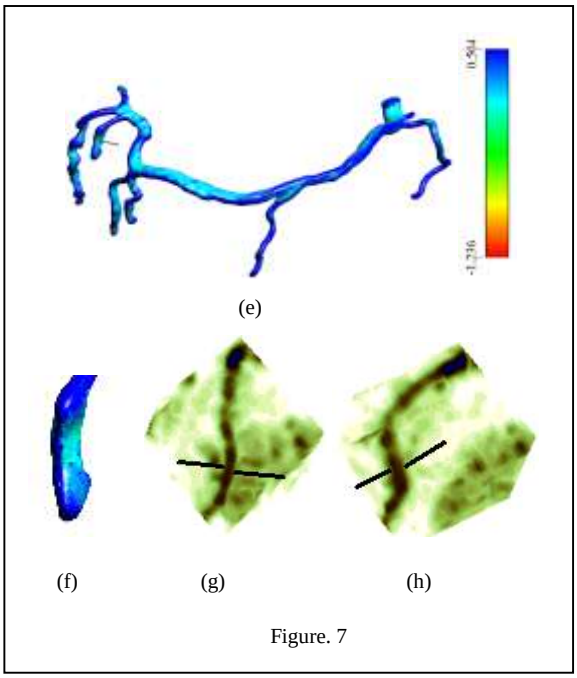
Figure. 7



(i) (j)

(k) (l)

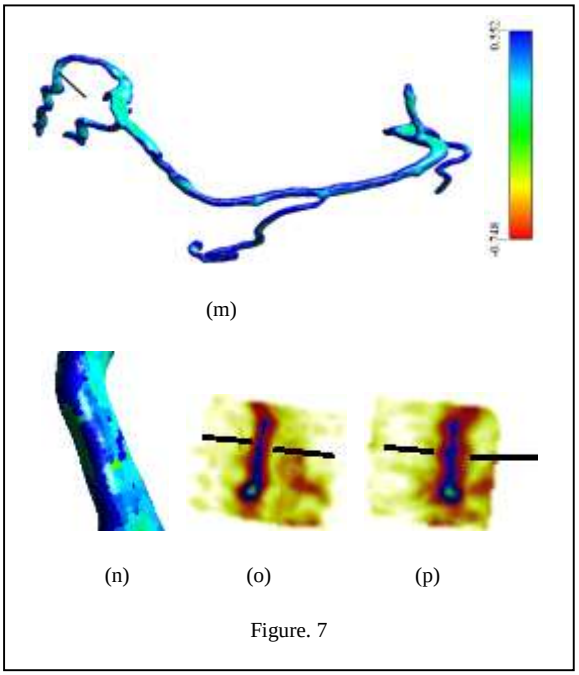
Figure. 7



(e)

(f) (g) (h)

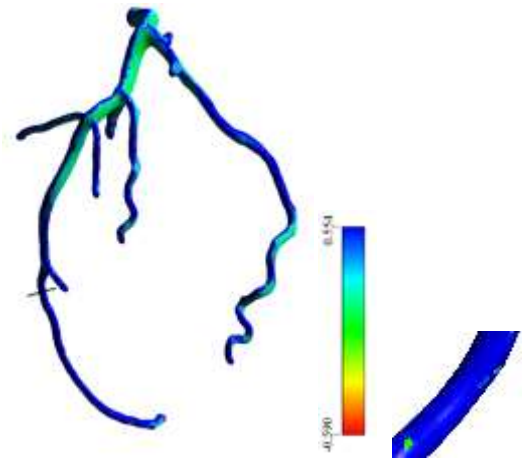
Figure. 7



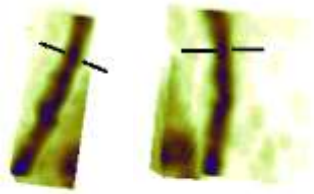
(m)

(n) (o) (p)

Figure. 7



(q) (r)



(s) (t)

Figure. 7 maximum principle curvatures mapped on the triangle meshed surface of (a) left coronary artery of dataset 1, (e) right coronary artery of dataset 1, (i) left coronary artery of dataset 2, (m) right coronary artery of dataset 2; (b), (f), (j), (n) large view of the artery segment located with line segment in (a), (e), (i), (m), respectively; (c), (g), (k), (o) small radius from the MIP image in one side view of a slab volume containing the artery segment in (b), (f), (j), (n), respectively; (d), (h), (l), (p) large radius from MIP image in another side view to the same slab volume in (c), (g), (k), (o) respectively.

another view of maximum principle curvatures mapped on the triangle meshed surface of (q) left coronary artery of dataset 1, (r) large view of the artery segment located with line segment in (q), (s) and (t) are MIP images in two sides views of the slab volume containing the artery segment located with the line segment in (q).

5. REFERENCES

[1] Geoffrey D. Rubin, Jonathon Leipsic, U. Joseph Schoef, Dominik Fleischmann, Sandy Napel, CT Angiography after 20 Years: A Transformation in Cardiovascular Disease Characterization Continues to Advance, Radiology, Vol 271: 3-June 2014.

[2] Paul S. Calhoun, Brian S. Kuszyk, David G. Heath, Jennifer C. Carley, Elliot K. Fishman, Spiral CT Data: Theory and Method, RadioGraphics, 19, 1999, pp745-764.

[3] Katja Bühler, Petr Felkel, Alexandra La Cruz, Geometric methods for vessel visualization and quantification – a survey, *Geometric modelling for scientific visualization*, G. Brunnet, B. Hamann and H. Müller (eds.), Kluwer Academic Publishers, 2004, pp. 399-420.

[4] Paul Schoenhagen, Arthur Stillman, Sandra S. Halliburton, Richard D White, CT of the Heart: Principles, Advances, Clinical Uses, Cleveland Clinic Journal of Medicine, 2, 2005.

[5] Jorge Larrey-Ruiz, Juan Morales-Sanchez, Maria C Bastida-Jumilla, Rosa M Menchon-Lara, Rafael Verdu-Monedero and Jose L Sancho-Gomez, Automatic Image-based Segmentation of the Heart from CT Scans, EURASIP Journal on Image and Video Processing, 52, 2014.

[6] Ali Hassan, Sarfraz Ahmed Nazir, Hatem Alkadhi, Technical Challenges of Coronary CT Angiography: Today and Tomorrow, European Journal of Radiology, Vol 79, 2011, 161-171.

[7] Victoria Yeh, Rine Nakanishi, and Matthew J. Budoff, Coronary Artery Disease Progression: Insights from Cardiac CT, Current Cardiovascular Imaging Reports, 8(7), 2015.

[8] Alan Conrad bovik, Handbook of Image and Video Processing, Elesvier Science, 2005.

[9] Gregory T. Wilson, Prabhakaran Gopalakrishnan, Tahir Tak, noninvasive Cardiac Imaging with Computed Tomography, Clinical medicine & Research, Vol 5, No. 3, 2007: 165-171

[10] T. Deschamps, P. Schwartz, D. Trebotich, P. Colella, D. Saloner, R. Malladi, Vessel segmentation and blood flow simulation using level-sets and embedded boundary methods, *International Congress Series*, 2004, 1268, pp. 75-80.

[11] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, Three dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images, *Medical Image Analysis*, 2, 1998, pp. 143-168.

[12] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale Vessel Enhancement Filtering, *MICCAI'98, Lecture Notes in Computer Science*, Vol 1496, pp130-137.

[13] Jayaram K. Udupa, Hsiu-Mei Hung, and Keh-Shih Chuang, Surface and Volume Rendering in Three-Dimensional Imaging: A Comparison, Journal of Digital Imaging, Vol 4, No 3 (August), 1991: pp 159-168

[14] Shawn Lankton, Arthur Stillman, Paolo Raggi and Allen Tannenbaum, Soft plaque detection and automatic vessel detection, *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop: Probabilistic Models for Medical Image Analysis*, 2009, pp. 25-33.

[15] Felix Renard, Yongyi Yang, Coronary artery extraction and analysis for detection of soft plaques in MDTC images, IEEE International Conference on Imaging Processing (ICIP) 2008, pp. 2248.

[16] A. Hennemuth, T. Boskamp, D. Fritz, One-Click Coronary Tree Segmentation in CT Angiographic Images, International Congress Series, Vol. 1281, 2005, pp. 317-321.

[17] Ola Friman, Milo Hindennach, Caroline Kuhnel, Heinz-Otto Peitgen, Multiple Hypothesis Template Tracking of Small 3D Vessel Structures, Medical Image Analysis, 14 (2010) 160-171.

[18] Michel Schaap, Lisan Neefjes, Coert Metz, Alina van der Giessen, Annick Weustink, Nico Mollet, Jolanda Wentzel, Theo van Walum and Wiro Niessen, Coronary Lumen Segmentation Using Graph Cuts and Robust Kernel Regression, The Series Lecture Notes in Computer Science: Information Processing in Medical Imaging, Vol. 5636, pp 528-539.

[19] Rahil Shahzad, Hortense Kirisli, Coert Metz, Hui Tang, Michiel Schaap, Lucas van Vliet, Wiro Niessen, Theo van Walsum, Automatic Segmentation, Detection and Quantification of Coronary Artery Stenoses on CTA, International Journal Cardiovascular Imaging, 29, 2013, pp 1847-1859.

[20] James Foley, Andries Van Dam, Steven Feiner, John Hughes, *Computer Graphic: Principles and Practices*, 2nd Edition, Addision-Wesley Publishing Company, Inc, 1990.

[21] William E. Lorensen, Harvey E. Cline: Marching Cubes: A high resolution 3D surface construction algorithm, *Computer Graphics*, 21(4), 1987, pp. 163-169.

[22] Nicole E. Jensky, Michael H. Criqui, Michael C. Wright, Christina L. Wassel, Steven A. Brody, Matthew A. Allison, Blood Pressure and Vascular Calcification, Hypertension, 2010, 55(4), 990-997.

[23] Michelle L. Frost, Rodolpho Grella, Sandrine C. Millasseau, Ben-yu Jiang, Geeta Hampson, Ignac Fogelman, Phil J. Chowienczyk, Relationship of Calcification of Atherosclerotic Plaque and Arterial Stiffness to Bone Mineral Density and Osteoprotegerin in Postmenopausal Women Referred for Osteoporosis Screening, Calcified Tissue International, 2008, 83, 112-120.

[24] Joshua D. Hutcheson, Natalis Maldonado, and Elena Aikawa, Small Entities with Large Impact: Microcalcifications and Atherosclerotic Plaque Vulnerability, Current Opinion in Lipidology, Vol. 25, October, 2014.

[25] Glaber S, Oeltze S, Hennemuth A, Kubisch C, Mahnken A,

Wilhelmsen S, et al. Automatic Transfer Function Specification for Visual Emphasis of Coronary Artery Plaque, *Computer Graphics Forum*. 2010; 29(1): 191-201

[26] Tobias Boskamp, Daniel Rinck, Florian Link, Bernd Kummerlen, Mildeberger, New Vessel Analysis Tool for Morphometric Quantification and Visualization of Vessels in CT and MR Imaging Data Sets, Radio Graphics, 24, 2004, 287-297.

[27] Rahil Shahzad, Hortense Kirisli, Coert Metz, Hui Tang, Michiel Schaap, Lucas van Vliet, Wiro Niessen, Theo van Walsum, Automatic Segmentation, Detection and Quantification of Coronary Artery Stenoses on CTA, International Journal Cardiovascular Imaging, 29 (2013), pp 1847-1859.

[28] Gordon Kindlmann, Ross Whitaker, Tolga Tasdizen and Torsten Moller, Curvature-based transfer functions for direct volume rendering: methods and applications, *Proceedings of the 14th IEEE Visualization* 2003, pp. 513-520.

[29] Barrett O Neil, Elementary Differential Geometry, Revised 2nd Edition, Elsevier Academic Press Publications, 2006.

[30] Henk A. Marquering, Jouke Dijkstra Patrick J.H. de Koning, Berend C. Stoel & Johan H.C. Reiber, Towards quantitative analysis of coronary CTA, *The International Journal of Cardiovascular Imaging*, 2005, 21, pp.73-84.

[31] M.J. van Gils, D. Vukadinovic, A.C. van Dijk, D.W.J. Dippel, W.J. Niessen, A. van der Lugt, Carotid Atherosclerotic Plaque Progression and Change in Plaque Composition Over Time: A 5-Year Follow-Up Study Using Serial CT Angiography, American Journal of Neuroradiology, Vol 33, August 2012.

[32] P.J. de Feyter, P.W. Serruys, K. Nieman, N. Mollet, F. Cademartiri, R.J. van Geuns, C. Slager, A.F.W. van der Steen, R. Krams, J.A. Schaar, P. Wielopolski, P.M.T. Pattynama, A. Arampatzis, A. van der Lugt, E. Regar, J. Ligthart, P.Smits, Imaging of Coronary Atherosclerosis and Identification of the Vulnerable Plaque, Netherlands Hearts Journal, Vol 11, No. 9, Septmber 2003.

[33] Simone Balocco, Carlo Gatta, Marina Alberti, Xavier Carrillo, Juan Rigla, Petia Radeva, Relation between Plaque Type, Plaque Thickness, Blood Shear Stress, and Plaque Stress in Coronary Arteries Assessed by X-ray Angiography and Intravascular Ultrasound, Medical Physics, 39(12), December 2012.

[34] Jonathon Leipsic, Suhny Abbara, Stephan Achenbach, Ricardo Cury, James P. Earls, GB John Mancini, Koen Nieman, Gianluca Pontone, Gilbert L. Raff, SCCT guidelines for the interpretation and reporting of the coronary CT angiography: A report of the Society of Cardiovascular Computed Tomography Guildelines Committee, *Journal of Cardiovascular Computed Tomography* 8, 342-358 (2014)

International Journal of Computer Applications Technology and Research
Volume 5–Issue 9, 612-618, 2016, ISSN:-2319–8656

A Review of Agile Software Effort Estimation Methods

Samson Wanjala Munialo.
Department of Information Technology
Meru University of Science and Technology
Meru - Kenya

Geoffrey Muchiri Muketha
Department of Information Technology
Murang'a University College
Murang'a - Kenya

Abstract: Software cost estimation is an essential aspect of software project management and therefore the success or failure of a software project depends on accuracy in estimating effort, time and cost. Software cost estimation is a scientific activity that requires knowledge of a number of relevant attributes that will determine which estimation method to use in a given situation. Over the years various studies were done to evaluate software effort estimation methods however due to introduction of new software development methods, the reviews have not captured new software development methods. Agile software development method is one of the recent popular methods that were not taken into account in previous cost estimation reviews. The main aim of this paper is to review existing software effort estimation methods exhaustively by exploring estimation methods suitable for new software development methods.

Keywords: Lines of codes, Cost constructive model, Function point, Agile, software effort estimation.

1. INTRODUCTION

Demand for more functionality, higher reliability and higher performance has resulted to higher competitiveness among software developers. To stay competitive, software developers need to deliver software products on time, within the budget and to the agreed level of quality. Most projects fail due to planning issues such as cost, time and requirements specifications. A study on software projects in 2012 by Standish shows that 43% of projects were challenged and 18% failed due to over budget, late delivery and less than required features or functions [1]. This illustrates the necessity for reliable software development method and software cost estimation method.

For faster and quality delivery, software vendors are moving from structured development methods where requirements are well known in advance to agile development which welcomes customer changing requirements at later stages of software development [2]. The shift from traditional development methods to agile is due to the high cost of affecting changes request by users at later stages of software development. Agile encourage changes, therefore decreasing the cost of change and reduce the overall development cost. Agile make this possible as a result of simple design, collective ownership, continuous testing and short releases cycles.

Many estimating methods have been proposed since 1950's and many studies have evaluated their effectiveness. The most popular traditional cost estimation methods are Expert judgment, Analogy, Wideband Delphi, Source lines of codes, Function points [3], Object points and Cost constructive model (COCOMO) [4] [5]. However, they are not effective when dealing with agile software estimation. Estimating agile software is a problem due to varying requirements and incremental development. This prompted the introduction of cost estimation methods such as planning poker [6] in 2003 which is one of the most popular agile estimation methods. However, planning poker depends on expert experience on previous projects and its estimates are specific to the team; another team may estimate different story point for the same project

Recently other methods such as Bayesian Belief Network, AgileMOW[7] and Constructive Agile Estimation algorithm [8] were introduced to deal with uncertainty and iterative nature of agile software development. Estimation of effort and cost depends on accurate estimation of the software size which helps to predict the project scope. Apart from size, other indicators such as project complexity factors are considered when estimating effort. Therefore, a reliable software cost estimation method must include critical cost indicators for more accurate estimation. The main objective of this paper is to discuss existing software cost estimation methods including their features and situations where they are applicable. This paper is organized in 6 sections which include introduction, background, Traditional effort estimation methods, agile effort estimation methods, discussion and conclusion.

2. BACKGROUND

Software estimation is a critical component of software project management. Cost overruns increased from an average of 56% in 2004 to 59% in 2012 in sampled software projects while time overruns increased from 71% in 2010 to 74% in 2012 [1]. More accurate estimation helps software developers to gain profit and customers to be more satisfied. On the other hand, high costing can lead to lost profit by losing bidding while low costing can lead to cost overruns and poor quality of end product. Software estimation comprise of estimating software cost, size, effort and time required to develop the software [9]. Software developers require an effective software cost estimation model to facilitate project planning and eventually successful implementation of a software project.

Agile software development method is one method that provides a challenge to existing software cost estimation techniques. Agile software development is based on iterative development where requirements evolve through collaborations. Scope is continuously adjusted throughout the project and new tasks are discovered [10]. It emphasizes on working software, customer collaboration, response to change on demand and does not support well defined requirements like the traditional waterfall method. All these challenges make most of the existing software cost estimation techniques to appear limited when dealing with agile software.

Software developers have had the interest of estimating accurately the cost of developing software products. The first methods were only based on software size using lines of codes or function points to estimate the cost. Currently other cost drivers such as process

factors and human factors have been included in estimation methods to improve on software estimation accuracy [7]. However, demand for new functionalities, quick delivery of software such as mobile applications established a need for new software development methods. Currently other features such software re-use, component based development, distributed systems and iterative development are common features in software engineering industry. Evolution in software engineering industry provided a challenge to software estimations researcher to come up with methods that will estimate more accurately.

3. TRADITIONAL COST ESTIMATION METHODS

Cost estimating models are classified as non-algorithmic and algorithmic. Non-algorithmic methods estimation relies on experts who have experience on similar previous projects while algorithmic methods use parametric in their estimation.

3.1 Non-Algorithmic estimation methods

Most non-algorithmic cost estimation techniques are based on analytical comparison with previous similar projects and expert experience [10]. Most popular methods in terms of recent publications in this group are expert judgment, Analogy, top-down, bottom-up, price-to-win and Wideband Delphi.

3.1.1 Expert judgment

Expert Judgment technique is the most frequently applied cost estimation method where experts are responsible for estimating the size and cost of a software. This method is based on the project manager experience in similar software projects. Expert cost estimation method is helpful when there is limitation in finding data and gathering requirements [10] [11] [12]. Expert judgment is prone to human errors and biasness. Its success is based on expert judgment that is expert experience may differ from one expert resulting to varying estimates on the same type of project. However, it is helpful in small and medium sized software project and when the development teams and software attributes have not experienced significant changes as compared to previous projects.

3.1.2 Analogy technique

Analogy technique estimation is done according to the actual cost of one or more completed projects that are similar to the new project to be estimated [8][10] [11]. Estimation can be done at the total project level or at sub system level. The strength of estimation by analogy is that the estimate is based on actual project experience and estimation can be done in the absence of an expert. However, it does not take into consideration the extent of other relevant cost factors in the previous project such as the environment and functions which may differ with new project cost factors [13]. In addition, a lot of past information about past projects is required whereas in some situations there may be no similar projects developed in the past to compare with.

3.1.3 Price-to-win, Bottom-up and Top-up

Price-to-win estimation method is based on customer budget instead of software parameters or features. Example is when a customer is willing to pay for 6 persons-month and the project estimate is 8 persons-month then estimation is done as per the customer ability to pay. This may cause delays and force developers to work overtime [13]. Price-to-win method helps in getting the contract but it generally causes cost and time overruns.

Bottom-up estimation method estimates by separating each software component then summed to give the overall estimate for the product. It is possible only when the requirements and design of the system are known at an early stage of software development [11] [14]. While top-down method established an overall estimate for the project then the system is sub-divided into its functional components which are then estimated based on the overall estimate [13] [14]. The design and requirements must be well defined to partition software to its component.

3.1.4 Wideband Delphi

Wideband Delphi method is a cost estimation technique where effort and cost are estimated centered on team consensus. It is done by getting advices from experts who have extensive experiences in similar projects. Wideband Delphi technique was introduced by Barry Boehm and John Farquher in 1970s. It uses work breakdown structure as the basis for estimating project size, effort and cost [12] [15]. This method emphasizes on consultations, communication and interaction among participants.

Participants include customer representatives and technical team members that will be involved in development of the software product. Each member estimates for each task and identify changes and missing assumptions in work breakdown structure. Members with high or low estimates are asked to justify, and then members revise the estimates. The cycle repeats until when estimators agree on the estimates. The coordinator collects estimates from team members and assembles the tasks and estimates into a single final task list.

Wideband Delphi depends on team members experience and agreement among members and thus it is not appropriate method when applied to a software project that is unfamiliar to members [14] [15]. Furthermore, it is a preferred method when requirements are well defined and therefore, cannot work for software development methodologies where requirements are not clear. However, it encourages collaboration among estimators. Lastly, the technique is simple to apply and supports consensus-based estimates. Even though Wideband Delphi estimates are consensus-based, experts may be biased, optimistic or pessimistic in their estimation given that this method cannot be quantified.

3.2 Algorithmic software cost estimation methods

These models use a formula to calculate the software cost estimate [13]. They rely on a combination of related cost factors which are input to mathematical equation to do the estimation. Most common algorithmic software cost estimation methods includes Source line of codes (SLOC), Object points, Function-Point(FP)[3], Constructive Cost Model-I (COCOMO-I) [4] and Constructive Cost Model-II (COCOMO-II) [5].

3.2.1 Source line of codes (SLOC)

Source line of codes is a size metric that illustrates the number of program statements and data definition but does not include comments. SLOC is the earliest cost estimation method used to estimate the size of FORTRAN and assembly language which are line based programming languages. SLOC uses historical data of a previously completed project of the same size whose SLOC was computed before then compared with the actual one to estimate project size. The size estimate is eventually used to estimate the project scope, effort and cost [10]. SLOC is dependent on the programming language and therefore cannot compare different

programming language lines of codes. Source line of codes cannot estimate the size of non-procedural languages and software complexity is not taken into consideration when estimating size.

3.2.2 Function Point Analysis

Albrecht's introduced Function point analysis method in 1983 [3] which had better estimation than source lines of codes. Function point is a size metric that quantifies the size and complexity of a software system with regard to functions that the system will deliver. Function count is arrived at by counting the basic software components which include external inputs, external outputs, external inquiries, logical internal files and external interfaces. Each of the function is weighed by complexity factor ranging from low, average to high [3] [11] [14]. Each function component is multiplied with a respective complexity level then summed up to give Function Count (FC).

Function point can be applied at requirement specification or design phase of system development [14]. Furthermore, function point is independent of language or methodologies used in software development [3]. Lastly, Non-technical user can easily understand the method. However, Function point cannot be used in situation where requirements are not clear such as in agile software development.

3.2.3 Object Point

It estimates the size of software based on number and complexity of objects [11] [17]. The objects are screens, reports and 3GL components. The steps for estimation effort using object point include: counting the number of objects, classification of objects (simple, medium, average), weight objects with regard to difficulty as shown in table 1.

Table 1: Classification of objects weight

Object type	Simple	Medium	Difficult
Screen	1	2	3
Report	2	5	8
3 GL components			10

Object point is determined by adding all the weights of object instances to get object point count.
Estimate percentage re-use then compute the overall object points (NOP) where, NOP = (Object Point) * (100-% reuse)/100

Furthermore, developers' productivity is weighted from low to highest then effort is estimated by dividing net object point by productivity [11]. It is easy to apply object point method at any stage of software development but on the other hand requirements must be well defined. Object point only considered 3GL and 4GL factors and thus cannot apply to current programming languages.

3.2.4 Constructive Cost Model (COCOMO)

COCOMO models were proposed by Barry Boehm [4]. These methods use parameters which were derived from previous experiences about software projects for estimation. Due to COCOMO methods popularity various studies have extended COCOMO framework to develop cost estimation methods with an aim of improving software estimation accuracy. The 4 COCOMO methods are simple COCOMO, Intermediate COCOMO, Detailed COCOMO and COCOMO II.

Basic COCOMO computes software effort and cost as a function of program size expressed in thousands lines of codes (KLOC) using the formula:

$$\text{Effort} = a(\text{KLOC})^b$$

Where a and b are complexity factors which are assigned weights according to software project complexity as shown in table 2.

Table 2: Complexity factor weights

Model	A	B
Organic (Simple)	2.4	1.05
Semi-detached(Average)	3.0	1.15
Embedded (Complex)	3.6	1.20

With the advancement in software development methods and environment, basic COCOMO was not able to capture all relevant cost factors in its estimation. Therefore, intermediate COCOMO was released to include emerging software attributes in their computation of software estimates.
Intermediate COCOMO uses Kilo lines of codes as in basic COCOMO but it includes EAF (Effort adjustment factors) which includes subjective assessment of products, hardware, personnel and project attributes [5] [13]. Effort adjustment factors consider a set of four factors, with each factor having a number of attributes. The complexity factors are hardware, personnel, project and product with the following attributes.

- Hardware attributes: Run-time performance constraints, Execution time constraint, Memory constraints, Volatility of the virtual machine environment and Required turnabout time.
- Personnel attributes : Analyst capability, Software engineering capability, Applications experience, Virtual machine experience, Programming language experience
- Project attributes: Use of software tools, Application of software engineering methods and required development schedule.
- Product attributes: Required software reliability, Size of application database and Complexity of the product, required reusability.

Each of the 17 attributes is rated on a 6 point scale that ranges from very low to very high. Based on the rating, an effort multiplier is determined and the product of all effort multipliers results is an *effort adjustment factor* (EAF). Typical values for EAF range from 0.9 to 1.4 The intermediate COCOMO model takes the form EFFORT = a* (KLOC)b * EAF.

Another COCOMO version is detailed COCOMO which incorporates all characteristics of intermediate COCOMO on each step of software development process (Analysis, Design, coding and testing). The 17 attributes are used in each step to estimate software development effort [5] [10] [11] [13].

COCOMO-II was introduced in 1997 is an extension of intermediate COCOMO. It predicts the amount of effort based on Person-Month (PM) in the software projects [5][13]. It uses Thousands lines of code or function point as the size metrics and the number Effort adjustment factors attributes were increased by 5 to 22 attributes. The Usage of COCOMO II is very wide and its results usually are more accurate. The 5 additional effort adjustment factors are:

- Precedents'(PREC)- Previous experience of the organization
- Development Flexibility (FLEX) –Degree of flexibility in development process.
- Risk resolution (RESL)- Extent of risk analysis carried out.
- Team cohesion (TEM)- How well development team knows each other
- Process maturity (PMAT)- Process maturity of the organization

COCOMO II formula takes the same format as intermediate COCOMO formula of estimating effort [5] [10] [11].

All COCOMO methods capture a wide range of parameter when estimating the cost of a project. So far COCOMO methods are the most popular methods with clear results. The use of COCOMO requires clear and well defined requirements [16]. However, a lot of data is required to estimate effort and the model is presented as black box to the user. However, COCOMO methods are challenged when requirements are not clear and when the project is subject to user request changes at later stages of software development.

3.2.5 Other software cost estimation methods

In recent years researchers have attempted to introduce more cost and effort estimation techniques to improve on estimation accuracy. One of the methods is Bayesian Belief Network which estimate software effort by forecasting software cost when information about the past and present is incomplete, vague and uncertain [18]. It includes a network of probabilities that captures the probabilistic relationship between variables in historical data [11]. The advantage of this method is not being dependent on knowing exact historical data. On the other hand, it requires knowledge of related parameters of previous project to be used in estimation.

The other method is Neural Networks which is based on the principle of learning from examples. Neural network use back propagation trained feed forward network to estimate software development effort [17] [11]. The network is trained with a series of inputs from previous projects to predict the effort of the current project. Neural network provided a more accurate estimate compared to other methods but it depends on data from previous projects.

4. AGILE COST ESTIMATION METHODS

The emergence of agile methods has presented many opportunities and challenges. One of the challenges is estimating the effort of developing agile software. Although traditional methods are used to estimate effort for agile software, they provide inaccurate results. Agile is a popular development method as it emphasize on collaboration with customer, communication among developers, rapid delivery of software and change of requirements on demand [20] [21]. Popular agile methods are Extreme programming, scrum, crystal, Feature driven development and learn development.

Some of the challenges of estimating agile methods include work assigned to a team and not an individual, emphasis is on collective effort and work is quantified in terms of effort rather than time and changing requirements on demand. Various studies were done in recent years and have come up with cost estimation methods suited for agile with the most popular one being planning poker [6]. Planning poker is a non-algorithmic method and is simple to implement. Other agile estimation methods introduced so far are constructive agile estimation algorithm [8] and AgileMOW [7] although their accuracy has not yet been calibrated by other researchers.

4.1 Planning Poker

Planning poker is an estimation method that is based on collaboration and consensus among team members like Wideband Delphi technique. It was initially proposed by Greening in 2003 and popularized by Cohn in 2005 [6] for agile software development such as scrum. Planning poker session is done at the beginning of an iteration of agile development involving a team of developers from different disciplines.

Each member in the team is given a deck of planning poker cards with values preferably Fibonacci sequence (1, 2, 3, 5, 8, 20, 40, 100) representing story points or ideal days. The nonlinear sequences reflect less uncertainty with smaller units and greater uncertainty when dealing with greater units [6]. A story in agile development is a brief description of functionality as viewed by the user or product owner. Story points are a relative unit of measure used to estimate the story size by taking into account effort, complexity and risk [19]. On the other hand, ideal days estimate a story with regard to the number of days or time it will take to translate a story to a system function or feature.

When a story has been fully discussed, each member privately estimates a story by selecting a card to represent the estimate. All cards are revealed at the same time and if the estimates are the similar then it becomes the agreed estimate. If not, high and low estimates are justified and discussed further. Then each member selects a card after the discussion and cards are revealed again. The process is repeated until consensus is achieved [19]. Two main reasons why planning poker is an effective way of estimating agile software is that it involves a team of experts from different disciplines who collaborate and justify their estimations to come with better results as compared to one expert providing estimate especially when there is high uncertainty and missing information.

4.2 Constructive Agile Estimation Algorithm

Constructive agile estimation algorithm was introduced in 2009 [8]. The algorithm uses vital factors namely project domain, performance, configuration, data transaction, complex processing, ease of operation and security which are weighed then incorporated in the estimation.

Constructive Agile Estimation algorithm divides estimation process into two phases called early estimation and Iterative estimation. The purpose of early estimation is to identify the initial scope just enough to draw the initial budget. Iterative estimation is done at the start of an iteration to include new requirements. In both cases story point is used to estimate the size of a feature as described by the user. Vital factors are identified on the grade of low, medium and high using Fibonacci series then multiplied to story point to get the final estimate.

Constructive Agile estimation algorithm identified factors that are critical in determining software effort but in addition people factors are also important especially in agile where collaboration and teamwork is an important ingredient for successful completion of a software project but they are not included in this algorithm.

4.3 AgileMOW

AgileMOW was introduced to estimate the cost of developing web applications using agile methods [7]. This method uses both expert judgment and algorithm to estimate effort. AgileMOW uses people and environment attributes described in COCOMO II which are aligned to agile manifesto. Factors used in this method include communication skills, proximity of team, feedback, courage, management skills, technical ability, reliability, ease of use and early delivery. The method use web objects to estimate size of a web application and people factors are weighed. Effort expressed in person-month is computed by multiplying web application size and weighted people and environment factors.

On main advantage of AgileMOW is that it identifies factors that align to the principles of agile software which focuses on communication and interaction. However, it cannot estimate the cost of other software rather than web application. Lastly, the method only focused on people factors whereas other factors such as product and process factors are also important when estimating agile software effort.

5. DISCUSSION

None of software cost estimation method is better or worse than the other, each has its own strength and weakness which are complementary to each other. Furthermore, software estimation methods are specific to a specific type of project or development method or software to be developed [5] [15]. Estimation methods such as Function point analysis, Object point and COCOMO are suitable when developing software in which requirements are fully known upfront such waterfall method. In contrast, these methods are challenged when requirements keep on changing such as in agile which require an estimation method that adapt to changes in such as planning poker estimation method.

Different situations and development environment determine the appropriate software cost method to be used. There are situations where accuracy in estimation is critical then a more accurate method should be employed, in other instance, winning a contract is important therefore, price-to-win becomes the most appropriate method [11]. Furthermore, small projects can easily be estimated using expert judgment but when the project becomes larger it requires more technical estimation method such as analogy and COCOMO. In addition, availability of data from previous project provides an opportunity to use analogy estimation method.

Several cost drivers should be considered to estimate software effort and cost. The most common cost driver among all estimation methods is the software size. Effort and cost can be estimated directly upon estimating the software size using one of the software size metrics such as source lines of codes, function point and object point. Agile size estimation is done using story point. Size is also used together with other factors to estimate software development effort when using most of algorithmic estimation methods. Therefore, software project managers must understand the key attributes in a project to identify an estimation method that will estimate accurately.

Each effort and cost estimation method has strengths and weaknesses based on the capabilities of the method. Table 4 shows a summary comparison of popular cost estimation methods.

Table 4: Comparison of software effort estimation method

Method	Strength	Weakness
COCOMO	- Clear results - Independent on programming language	- Much data required - Requirements must be clear. - Not adopted to changes in requirements
Function point	- Clear results - Independent on programming language	- Requirements must be clear. - Not adopted to changes in requirements
Expert	- Less data required - Adopt to special projects	- Its success depend on the expert
Analogy	- Based on similar project experience - More accurate	- Information about past projects is required - Historical data may not be accurate
Price-to-win	- Gets contract	- High overruns
Top-down	- Faster to implement System level focus - Minimal project details required	- Less stable - Less detailed
Bottom-up	- Based on detailed analysis - Support project tracking	- Difficult to estimate early in the life cycle - Time consuming
Wideband Delphi	- Reduced biasness by involving a team of experts	-Its success depend on the expert -Not adopted to changes in requirements
Planning Poker	- Adopt to changes in requirements - Reduced biasness by involving a team of experts	- Its success depend on team of experts - Estimation is relative to a team.

6. CONCLUSION

This paper provided a comprehensive overview of existing software cost estimation models describing their strengths and limitations. It is important for the software project manager to understand key factors relevant in estimating the cost of software and situations where an estimation method will be appropriate. No existing model can estimate the cost of software development with a high degree of accuracy, therefore the study of software cost estimation is necessary to improve on estimation accuracy.

With the emergence of new software development methods and techniques, future work will be to identify key estimation indicators in new software development methods and devise new cost estimation method.

7. REFERENCES

[1] The Standish group, 2013, Chaos Manifesto: Think big, Act small, The Standish Group Intenational.

[2] Coelho, E.,Basu, A.,2012, Effort Estimation in Agile Software Development using Story Points, International Journal of Applied Information Systems.

[3] Albrecht, A.J., Gaffney, G.E., 1983, Software Function, Source lines of Codes, and Development Effort Prediction: A Software Science Validation, IEEE Trans Software Engineering.

[4] Boehm, 1981, Software Engineering Economics, Prentice Hall.

[5] Boehm, B.W. et al, 2000, Software Cost Estimation with COCOMO, Prentice-Hall.

[6] Cohn, M.,2006, Agile Estimating and Planning, Pearson Education

[7] Litoriya, R., Kothari, A.,2013, An Efficient Approach for Agile Web Based Project Estimation: AgileMOW, International Journal of Computer Science and Computer applications.

[8] Bhalerao, S., Ingle,M.,2009, Incorporating Vital factors in agile estimation through algorithmic method, International Journal of Computer Science and Computer applications.

[9] Ziauddin, Tipu, S.K., Zia, S., 2012, An Effort Estimation Model For Agile Software Development, Advanced Computer Science and Its Applications.

[10] Khatibi, V., Jawawi, D.N., 2010, Software Estimation Methods: A review, Journal of Emerging Trends in Computing and Information Sciences.

[11] Borade, G, J.,Khalker, R,V.,2013, Software project effort and cost estimation techniques, International Journal of Advanced Research in Computer Science and Software Engineering.

[12] Gandomani, T.,Wei, T., Binhamid, K., 2014, Software Cost Estimation Using Expert Estimates, Wideband Delphi and Planning Poker Technique, International Journal of Software Engineering and its applications.

[13] Kumari, S., Pushkar, S., 2013, Performance Analysis of software cost Estimation methods: A Review, International Journal of Advanced Research in Computer Science and Software Engineering.

[14] Sharma, N., Bajpai, A., Litoriya, R., 2012, Software Effort Estimation, International Journal of Computer Science and Applications.

[15] Stellman, A., Greene, J.,2005, Applied Software Project management, O'Relly Media.

[16] Basha, S.,Dhavachelvan, P.,2010, Analysis of Empirical Software Effort Estimation Model, International Journal of Computer Science and Information Security.

[17] Bogdan, S., 2003, Software Development Cost Estimation Methods and Research trands", Computer Science.

[18] Angyan Y., Charlottesville, 2003, A Bayasian Belief Network approach to certifying Reliability of COTS software systems", Annual Reliability and maintainability Symposium, IEEE.

[19] Calefato, F., Lanubile, F., A Planning Poker Tool for Supporting Estimation in Distributed Agile Development, The Sixth International Conference on Software Engineering Advances

[20] Cao, L., 2008, Estimating Agile Software Project Effort: An empirical study, Association of Information Systems AIS Electronic Library(AISeL), Americas Conference on Information Systems.

[21] Schmietetendorf, A., Kunz, M., Dumke, R, 2008, Effort Estimation for Agile Software Development Projects, Proceedings 5th Software Measurement European Forum, Milan