

STR/mtDNA Visualization Profiler Analysis Using Jmol

Noor Maizura Mohamad Noor, Noraida Hj Ali, and Mohd Iqbal Hakim Harun
Universiti Malaysia Terengganu, Malaysia
Email: {maizura, aida}@umt.edu.my, GSK1599@pps.umt.edu.my

Abstract—Visualization process is a series of transformations to convert raw simulated data into a displayable image. The human brain can analyze pictures and patterns better than computers. In visualizing mtDNA molecule, Jmol modeling being is used. Jmol is a Java-based molecular modeling application that can be used either in a stand-alone mode, or as an embedded object in web pages. With Jmol modeling, Human mitochondria can be viewed to see the buildup of the molecule where it will contain multiple copies of a circular genome made up of double-stranded DNA (mtDNA). In criminal investigation, the presentation data to the juries remains a critical feature in the forensic investigation. In a DNA profile, it consists of measurements of the genotype at a number of marker loci. The STR/mtDNA profiler will portray a DNA sequence using a visualization technique where it is relevant not only to avoid or detect mistakes in the production and interpretation of visualization, but also helps to reduce mistakes in the production of DNA sequence and the interpretation of it.

Index Terms—algorithm, e-Learning, jmol, mtDNA, visualization

I. INTRODUCTION

In the last decade, visualization is an efficient way of transferring a large amount of information from a database into understandable and meaningful picture for individual understanding. Visualization process is a series of transformations to convert raw simulated data into a displayable image. The human brain can analyze pictures and patterns better than the computer. If there are no tools and applications specifically designed for the raw to be discussed, the visualization applications can create rich opportunities for users to engage in discussions about the data being displayed [1]. Scientists can usually recognize incorrect results in graphical depictions of the output of their algorithms. To find the source of errors, we need a way to apply this same visual understanding to the internal logic of the algorithms.

An algorithm for animation of a data allows the data to be visualized for the behavior of an algorithm. Although algorithm to animate a data has mostly been used in computer science, potential applications in other scientific domains also need for example mathematics, economics, and many other things.

An algorithm that creates a visualization of both the data and the operations of the algorithm. The algorithm maps the data into an image, which then gets animated based on the operations between two succeeding states in the execution of the algorithm. Animating an algorithm allows for better understanding of the inner workings of the data. Furthermore, it makes apparent shortcomings and advantages of the data and allowing it for further optimization.

The scientific forensic algorithms method of gathering and examining DNA information also is not left behind by using DNA testing, where it has and even many criminal defense lawyers to tell about the past which is then used in a court of law, on the idea that DNA tests provide a unique and infallible identification. DNA evidence has sent thousands of people to prison and, in recent years, have played a vital role in exonerating men who were falsely convicted, where the information can get from the person DNA or DNA sample.

Deoxyribonucleic acid or DNA stands for molecules like a double-stranded string contains information, blueprint, barcode or instruction for each cell in the human body or plat that will be inherited from both of children parents. DNA is present in every cell of our body, and without we realize that we also leave cells behind anywhere we go. Flakes of skin, drops of blood, hair, and saliva all contain DNA that can be used to identify us [2]. If the DNA that being found is being degraded the DNA can still be used in the court by extracting the mtDNA.

The mitochondria contain small amounts of DNA. This genetic material is known as mitochondrial DNA or mtDNA. Most of the mtDNA sequences are too long to be wholly displayed on the computer monitor, and it is very difficult to reveal some peculiarities or characteristics directly from the text string that represents DNA [3]. This is because the visualizations designed to support communication and collaborative analysis of the data. Most visualization relies on interaction and is far less effective in the screenshot or printed form [4]. A good way to visualize an mtDNA sequence would help in estimating nucleotide content, and in revealing repeated or missing subsequences, conserved regions [5].

II. THE BASIC OF FORENSIC PROCEDURE

Personal identification relies on identifiable characteristics such as biological (DNA, mtDNA, blood,

saliva, etc.), physiological (fingerprints, eye irises and retinas, hand palms and geometry and facial geometry), behavioral (dynamic signature, gait, keystroke dynamics, and lip motion) and on mixture of physiological and dynamical characteristics such as the voice.

mtDNA has become the most important personal identification characteristic because all genetic differences, whether being expressed regions of mtDNA (genes) or some segments of mtDNA with no known coding function but whose pattern of inheritance can be monitored can be used as markers. Any two humans are more than 99% identical in their DNA sequences, still have millions of genetic differences, making them different in their risk of getting certain diseases and response to environmental factors.

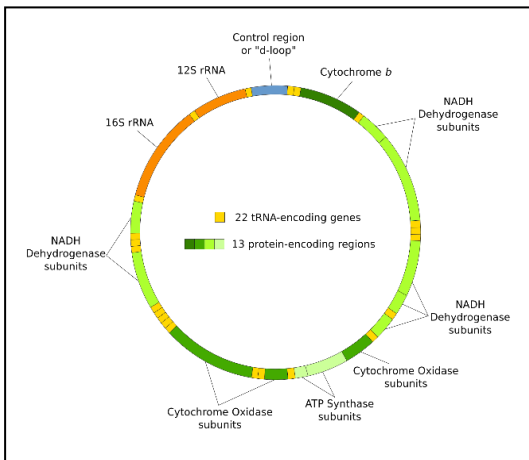


Figure 1. Human mitochondrial DNA.

Fig. 1 shows mitochondrial DNA can be assessed as the smallest chromosome coding for 37 genes and containing approximately 16,600 base pairs. Human mitochondrial DNA was the first significant part of the human genome to be sequenced. In most species, including humans, mtDNA is inherited solely from the mother. Mechanisms for this include simple dilution, degradation of sperm mtDNA in the fertilized egg, and, at least in a few organisms, failure of sperm mtDNA to enter the egg. Whatever the mechanism, this single parent (uniparental) pattern of mtDNA inheritance is found in any human, animals and also in plant.

Repetitions occur at most of sequence length scales, number and dispersion [1]. Examples of such repetitions are homo- and di-nucleotide repeats (microsatellites), and families of interspersed, mobile elements hundreds of base pairs long such as the ALU sequences. There are more than one million ALU sequences in the human genome, each 300 bases long, which are able to copy themselves in other parts of the genome, generating mutations.

III. OBJECTIVE

The main goal in the development of STR/mtDNA profiler is to combine DNA testing sample of STR and mtDNA only at one place and to visualize for understanding of the *practitioner* especially students. The

final result of the testing is expected to be fast and cost-effective compared to the normal DNA testing. The law enforcement unit may exonerate innocent convicted people. There are a number of ‘cold cases’. Therefore, STR/mtDNA profiler used as a standard marker of loci for determining the differentiation between sample at the crime scene or from contributed samples. The profiler mtDNA testing is one of the forensic techniques championed by both defense lawyers and prosecutors. Development of STR/mtDNA profiler will contribute to reduce crime rates in Malaysia as one of the latest initiative under National Key Result Areas (NKRA).

IV. METHODOLOGY

Visualization is an efficient way of transferring a large amount of information from a database into understandable information to an individual's head. If there are no specifically designed for the raw to be discussed, the visualization applications can create rich opportunities for users to engage in discussions about the data being displayed. This is because the visualizations designed to support communication and collaborative analysis of the data [4]. Most visualization relies on interaction and is far less effective in the screenshot or printed form.

Most of the visualization data rely on interaction and are far less effective in the screenshot or printed form, but, nowadays there are many devices with multi touch gesture controls, and it is always-on with connectivity and fast processing power, that bring all the visualization like static graphs and pie charts are being brought to life through real-time data.

In our research to visualize mtDNA molecule, we used Jmol modeling. Jmol is a Java-based molecular modeling application that can be used either in a stand-alone mode, or as an embedded object in web pages. Over the past couple of years, Jmol has emerged as a viable replacement for two long-standing tools used for visualizing macromolecules on personal computers, Because it is Java based is essentially independent of the platform that it runs on, when used from a webpage, it requires only a Java enabled browser.

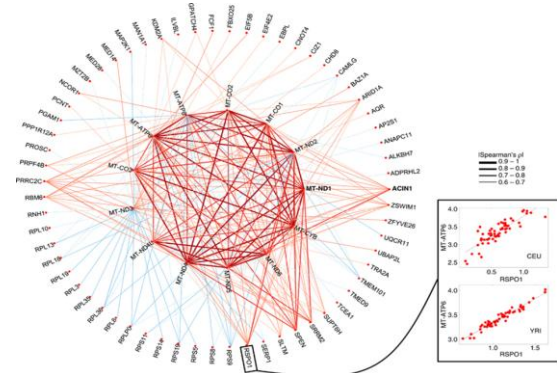


Figure 2. Co-varying expression network between mtDNA and nuclear genes.

Fig. 2 shows a co-varying expression network between mtDNA and nuclear genes. The consensus network

shared between CEU and YRI is plotted. Connections are placed between genes that are correlated at Spearman's $r=0.60$ in CEU and YRI.

A. Mitochondrial DNA

The human cell also contains DNA molecules outside the nucleus, known as mitochondrial DNA (mtDNA). mtDNA is maternally inherited, i.e. it is passed in identical form from mother to child, ignoring mutation. This makes mtDNA important for evolutionary genetics. But it is also significant for forensic identification: Two persons which are related through a maternal line will have (almost) identical mtDNA. As mentioned, mtDNA is maternally inherited and passed unchanged from mother to child. Similarly, the Y-chromosome is paternally inherited, i.e. passed from father to son in identical form. So two male individuals related through a paternal line will have identical Y-chromosomes.

The mitochondrion is the primary energy-generating organelle, harboring critical components of the electron transport chain for ATP synthesis through oxidative phosphorylation. In almost all eukaryotic cells, mitochondria have central roles in biosynthesis, homeostasis, and programmed cell death.[6]–[8].

Dysfunctional mitochondria have pleiotropic negative effects, giving rise to a large spectrum of defects that primarily affect tissues with high energy requirements such as the brain, heart, liver, skeletal muscles, kidney, and the endocrine and respiratory systems. [9]–[12]. To simultaneously estimate the expression of mtDNA and nuclear genes, the present study published RNA-seq data [13], [14] that included a large number of polyadenylated [1], [15]–[17] mitochondrial transcripts. The quantified between-individual variation in the expression of mtDNA genes in HapMap samples of European (CEU) and African (YRI) ancestry to evaluate this variation at the population level. The co-varying expression between mtDNA and nuclear genes, and explored various potential mechanisms that may underlie the association. Human mitochondria contain multiple copies of a circular genome made up of double-stranded DNA (mtDNA) that encodes proteins involved in cellular respiration. Transcript abundance of mtDNA-encoded genes varies between human individuals, yet the level of variation in the general population has not been systematically assessed.

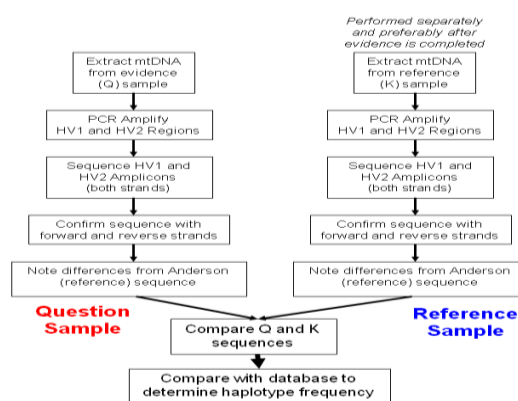


Figure 3. Process for evaluation of mtDNA samples.

Fig. 3 showed how the process is being done to evaluate the mtDNA between two different samples. In criminal investigation the presentation data to the juries remains a critical feature of forensic mtDNA usage. Emphasis on the non-unique status of the marker is important so that juries are not confused about the differences between the powerful statistics used for nuclear DNA and the more modest statistics possible with mtDNA results [1]. Recently no court decision has been overturned due to any misrepresentation during testimony about the strength of the statistical conclusion or a failure to represent the non-unique haploid mode of mtDNA inheritance.

B. Visualization

In [18] visualization is to form a mental vision, image, or picture of (something not visible or present to sight, or of an abstraction); to make visible to the mind or imagination. Modern biology, chemistry, biochemistry and crystallography education has benefitted from fast desktop computers and free and commercial software for the visualization of crystal and molecular structures. The software supports instructors to demonstrate, and student's practice, operations that would be nearly impossible to carry out by hand. But a simple technique of visualization, can transform students of all ages from passive to active readers; visualization can help students cross the boundary to improved comprehension. When visualizing data using a digital computer, the data must be in digital form. There are a great variety of data-sources. Digital data may be produced by the various experimental equipment.

Visualization is the graphical presentation, display of information, with the goal of providing the viewer with a qualitative understanding of the information contents. It is also the process of transforming objects, concepts, and numbers into a form that is visible to the human eyes. Data visualization is all about understanding ratios and relationships among numbers. Not about understanding individual numbers, but about understanding the patterns, trends, and relationships that exist in groups of numbers [2]. It is important to note that for a visualization to be effective, it must draw upon the knowledge base of the viewer. If the viewer does not possess the knowledge to understand the graphical entities and the relations between them the visualization does not achieve its goal. Visualization has many applications. For the most part they can be classified into two categories:-

- Data Exploration
- Communicating Information

Data Exploration is the practice of using visualization techniques to find unforeseen relationships between data points or sets of points in large databases. Once a relationship has been found the same visualization can be used to communicate that relate to others. For example, educational visualization where it will be using a simulation not as normal traditional teaching where the visualization were created on a computer to create an image of something so it can be taught about. This is very useful when teaching about a topic that is difficult to otherwise see, for example, atomic structure, because

atoms are far too small to be studied easily without expensive and difficult to use scientific equipment.

C. Jmol

Jmol is an open-source viewer for three-dimensional chemical structures, with features for chemicals, crystals, materials and biomolecules that does not require 3D acceleration plugins. Jmol returns a 3D representation of a molecule that may be used as a teaching tool [4], or for research e.g. in chemistry and biochemistry.

V. RESEARCH ARCHITECTURE

Fig. 4 shows that how the STR / mtDNA system architecture profiler that has been developed to matching between two DNA samples, one that been found from crime scene and the other one sample is will be match within the DNA database or a sample from the suspect.

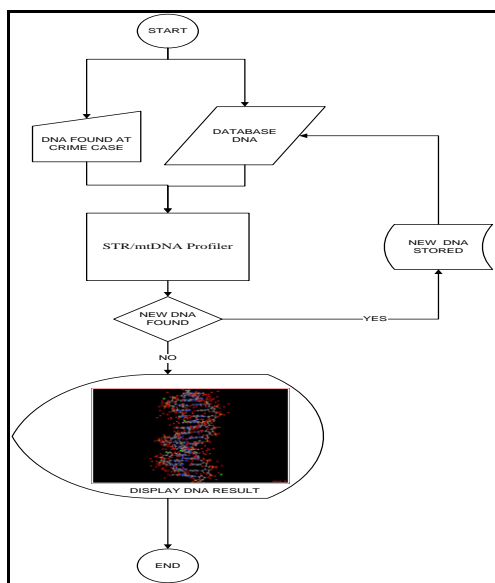


Figure 4. Research visualization algorithm.

This application disputes the paternity analysis whether the individual A is the father of individual B. STR/mtDNA profiler may help in solving criminal cases, if crime scene investigators found stain of evidence and they try to solve who that contributed to the stain evidence.

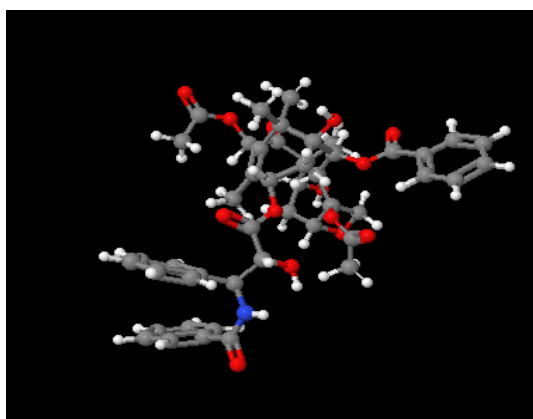


Figure 5. Full molecule visualize using jmol.

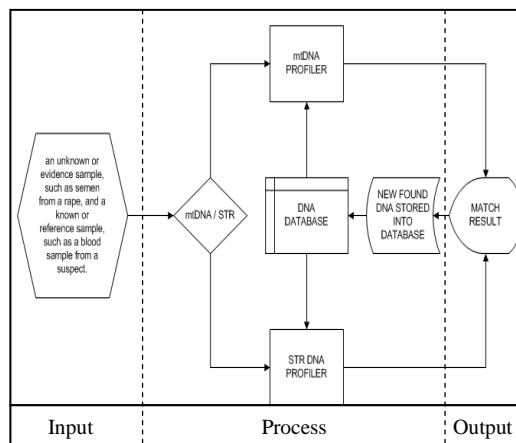


Figure 6. System architecture STR/mtDNA profiler.

Fig. 5 shows how jmol will produce final molecule that been convert from STR/mtDNA sequence. Typical DNA cases involve the comparison of two samples where an unknown or evidence sample, both can come from semen in rape cases, or it can come from a known or reference other than the semen, such as a blood sample from a suspect.

Fig. 6 shows the process flow on how the STR/mtDNA system architecture profiler matching between two DNA samples, one that had been found from crime scene and the other DNA database or samples of the dna from the suspect. The DNA profile that being found at a crime scene is very limited, and we can't easier to imagine that the suspects might be unrelated to the crime and that the suspect matches only by chance [19].

VI. IMPLEMENTATION OF DNA PROFILE AND GENOTYPES

The genotype of an individual at a given locus is the unordered pair of alleles at that locus. One cannot measure which allele originated from the mother and which from the father. The genotype is typically reported as (12, 14) or (A,B), so that the smallest is mentioned first

TABLE I. TWO PERSON SIMULATED MIXTURE—NUMBER OF ALLELES AT EACH LOCUS

Loci	No of Alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441
D8	0.011	0.194	0.521	0.274
D21	0.007	0.147	0.505	0.341
D18	0.003	0.095	0.472	0.430
D19	0.020	0.261	0.516	0.203
THO	0.016	0.271	0.547	0.166
FGA	0.003	0.116	0.500	0.381

Source : Buckleton *et al.*, Forensic Science International:Genetics 1(2007) 20-28.

Table I shows a DNA profile consisting of measurements of the genotype at a number of marker loci. Standard kits use 9 or 10 markers, but occasionally more markers are measured. Markers are generally chosen on different chromosomes, to avoid problems of linkage, i.e. dependence created in the process of meiosis, in the STR/mtDNA profiler as shown in figure 5 is the example of molecule that being visualizes using jmol. If the DNA profile consists of a combination of traits that figure to be extremely rare, the evidence is very strong that the suspect is the contributor.

Figure 7. Developed DNA matching form.

VII. DISCUSSION

It is generally believed that mtDNA gene expression is proportional to the number of mtDNA copies [19], [20] and the amount of mtDNA in a cell could provide a major regulatory point in mitochondrial activity [19], [21]. Functional analyses of correlated genes and modules confirmed biological processes previously known to be associated with mitochondrial activities, such as apoptosis [22]. These analyses also identified many genes involved in those biological processes previously unknown to be associated with mitochondrial activities [23].

VIII. CONCLUSION

A forensic science process through mtDNA analysis provides a lead to the suspects in criminal investigations. Faster result and effective decision making can be acquired, even though without partial complete DNA or DNA without nucleus or even hair without root – compared with DNA analysis using STR technique. Analysis of DNA data has become one of the main reliable results for victim identification either in crime investigation or even natural disasters. The STR/mtDNA profiler used the advancement of science and technology will continue to protect the harmony in communities. The STR/mtDNA profiler will portray a DNA sequence using a visualization technique where it is relevant not only to avoid or detect mistakes in the production and interpretation of visualization that helps to reduce mistakes in the production of DNA sequence and the interpretation of it. Visualizations of the DNA sequence

help authorities get more understanding on DNA analysis and therefore will assist in solving crime more effectively.

ACKNOWLEDGMENT

The authors would like to thank for continuous supports given by Datuk Jamshah Mustapa, Chief Officer of Criminal Investigation Department Terengganu, Royal Police Malaysia (RPM). This work was supported by a grant from Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education (MOHE) with vote number 59289.

REFERENCES

- [1] F. Piazza and P. Lio, "Statistical analysis of simple repeats in the human genome," *Physica A*, vol. 347, pp. 472-488, 2005.
- [2] K. Norrgard, "Forensics, DNA fingerprinting, and CODIS," *Nature Education*, vol. 1, no. 1, p. 35, 2008.
- [3] Mitochondria DNA: The Eve Gen.
- [4] F. B. Viçgas and M. Wattenberg, "Communication-minded visualization: A call to action," *IBM Systems Journal*, vol. 45, no. 4, p. 801, 2006.
- [5] A. Myllari, T. Salakoski, and A. Pasechnik, "On the visualization of the DNA sequence and its nucleotide content," *ACM SIGSAM Bulletin*, vol. 39, no. 4, pp. 131-135, 2005.
- [6] L. G. Biesecker, *et al.*, "DNA identifications after the 9/11 world trade center attack," *Science*, vol. 310, pp. 1122-1123, 2005.
- [7] E. Huffine, J. Crews, B. Kennedy, B. Kathryn, and Z. Asta, "Mass identification of persons missing from the break-up of the former Yugoslavia: Structure, function, and role of the international commission on missing persons," *Croat Med. J.*, vol. 42, no. 3, pp. 271-275, 2001.
- [8] P. Gill, "DNA as evidence—the technology of identification," *N. Engl. J. Med.*, vol. 352, no. 26, pp. 2669-2671, 2005.
- [9] H. D. Cash, J. W. Hoyle, and A. J. Sutton, "Development under extreme conditions: Forensic crime investigation in the wake of the world trade center disaster," in *Proc. Pac. Symp. Biocomput.*, 2003, pp. 638-653.
- [10] P. Estrela, A. G. Stewart, F. Yan, and P. Migliorato, "Field effect detection of biomolecular interactions," *Electrochimica Acta*, vol. 50, pp. 4995-5000, 2005.
- [11] P. Estrela, A. G. Stewart, P. Migliorato, and H. Maeda, "Label-free detection of DNA hybridization with Au/SiO₂/Si diodes and poly-Si TFTs," in *Proc. Technical Digest of 2004 IEDM International Electron Devices Meeting*, San Francisco CA: IEEE, 2004, pp. 1009-1012.
- [12] J. Califano, "Interview with amanda parrish," *Mitochondrial Genome Scan Finds Cancer Mutations in Saliva DNA Samples. Affymetrix Microarray Bull.*, vol. 1, pp. 17-19, 2005.
- [13] L. A. Foreman, A. F. M. Smith, and I. W. Evett, "Bayesian analysis of deoxyribonucleic acid profiling data in forensic identification applications," *Journal of the Royal Statistical Society: Series A*, vol. 160, pp. 429-469, 1997.
- [14] I. W. Evett and B. S. Weir, *Interpreting DNA Evidence 1998*, Sunderland: Sinauer, 1998.
- [15] R. Redon, *et al.*, "Global variation in copy number in the human genome," *Nature*, vol. 444, no. 7118, pp. 444-454, 2006.
- [16] D. Gagliardi, P. P. Stepien, R. J. Temperley, R. N. Lightowlers, and Z. M. A. Chrzanowska-Lightowlers, "Messenger RNA stability in mitochondria: Different means to an end," *Trends in Genetics*, vol. 20, pp. 260-267, 2004.
- [17] S. Slomovic, D. Laufer, D. Geiger, and G. Schuster, "Polyadenylation and degradation of human mitochondrial RNA: The prokaryotic past leaves its mark," *Mol. Cell. Biol.*, vol. 25, pp. 6427-6435, 2005.
- [18] J. Simpson and E. S. C. Weiner, *The Oxford English Dictionary*, Oxford: Clarendon Press, 1989.
- [19] S. Magazine. (August 2007) Data visualization: Modern approaches. Graphics [Online]. pp. 1-39, Available: http://www.science.smith.edu/dftwiki/images/9/96/DataVisualization_SmashingMagazine.pdf

- [20] Z. Vincent, *3D Structure Visualization and High Quality Imaging*, Chimera, 2008.
- [21] C. Demetrescu, I. Finocchi, G. F. Italiano, and S. Naher, "Visualization in Algorithm Engineering: Tools and Techniques", in *Experimental Algorithmics*, vol. 2547, R. Fleischer, B. Moret, and E. M. Schmidt, Eds., Berlin Heidelberg: Springer, 2002, pp. 24-50.
- [22] T. T. Elvins, "A survey of algorithms for volume visualization," *ACM Siggraph Computer Graphics*, vol. 26, no. 3, pp. 194-201, 1992.
- [23] W. C. Jeremy, "Maps as social constructions: Power, communication and visualization," *Progress in Human Geography*, vol. 25, no. 2, pp. 235-252, 2001.



Dr. Noor Maizura Mohamad Noor is Associate Profesor at Faculty of Science and Technology, Universiti Malaysia Terengganu (UMT). She received her PhD degree from The Manchester University, UK in 2005. Her recent research work focuses on improving organizational decision making practices through the use of technologies. This includes research interests in the design, development and evaluation of decision support systems for analyzing and improving decision

processes.

Her research interests also focus on the areas of computer science, intelligent decision support system, clinical decision support system, and information system. She is an editor of books, conferences proceedings and also a reviewer and editorial board of International Journals. Her contribution to scientific computing has been recognized by receiving several awards including the Best Scienties Award (2009

and 2013), Product and Research Innovation Award for International and National Level (2008 and 2009) of UMT. She is a member of ACM and IACSIT.



Dr. Noraida Haji Ali is a lecturer at the School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT). She obtained her B.Sc. and M.Sc. degrees in Computer Science from Universiti Kebangsaan Malaysia in 1995 and 1999, respectively. Five years later, she furthered her study leave for doctoral level at University Kebangsaan Malaysia, specialized in object-oriented modeling. Her current research interests focus on the software engineering area especially in development of systems, object-oriented modeling and formal method modeling. She is actively conducting her research in her field of interests through the supervision of undergraduate and postgraduate students. She has supervised more than 50 undergraduate students as a main supervisor and 3 postgraduate students (M.Sc) and 2 student PhD as a main supervisor and co-supervisor. She has received several research grants from the university either as leader and co-researcher.



Mohd Iqbal Hakim Bin Harun is currently a Master degree student in Computer Science Department, University Malaysia Terengganu. He received degree in Computer Science (Software Engineering) from University Malaysia Terengganu in 2012. His research interest is information system, decision support system.