Supporting Decision-making for Forensic DNA Analysis in Crime Investigation using Visualization

Noor Maizura Mohamad Noor[1], A. F. Ghazali[1], M. Y. M. Saman[1], and Zafarina, Z[2]
[1] Universiti Malaysia Terengganu, Terengganu, Malaysia
Email: {maizura@, gsk0909@pps., yazid@}umt.edu.my
[2] Universiti Sains Malaysia, Kelantan, Malaysia
Email: zafarina@kck.usm.my

***Abstract*—One of the challenging issue that scientists are facing is to have useful and effective data visualizations specifically for forensic DNA analysis in crime investigation. Concept of data visualization can then be employed into Bayesian network in order to interpret complex linking of data into interactive graphs for ease of analysis and accurate decision-making. In forensic DNA analysis, using visualization can make the activity of interpreting data and searching for details become easier and faster. The complexity of information in this field requires a good way of representation to analyze DNA profiles in crime investigation. In this paper, we show an implementation of visualizing forensic DNA analysis for suspect and unknown DNA profiles using a web prototype.**

***Index Terms*—visualization, decision support system, forensic, DNA, crime investigation, Bayesian network**

I. Introduction

Data visualization is a technique in computer science that means to help represent data in an effective way. Interactive and systematic analysis will help forensic scientists and law enforcement gain insight to make quicker and better decision. Visualization has contributed a lot in recent research works [1-4] due to its effectiveness [5-8]. When considering data sets with many features, the issues are which features to visualize and how to combine them in the visualization [2].

The question that should be answered is how do we make sense of the data? How do we harness this data in decision making processes? And how do we avoid from being overwhelmed? It is very important to find out how to visualize complex set of DNA profile data in an easier way for us to understand and retrieve any details.

A. Computerized DNA analysis

Deoxyribonucleic acid (DNA) analysis is a time-consuming and critical process. Decisions need to be made based on available choices. A large amount of forensic DNA data is difficult to interpret in one quick glance. Related works in forensic DNA analysis tools: softwares used for forensic DNA analysis was reviewed by [9] including DNAStat (2007), DNA-View (2008), EasyDNA (2008), EasyPate (2000), familias (2006), GenoProff (2008), Genotype (2008), Hugin (2004), PatCan (2003), Paternity Index (2007) and PatPCR (2003).

Similarities among them are: simple and short programs, but they can still be complex due to branching alternatives from different inputs and its data processing into results. None taking attention on visualization, hence visualization should be applied in forensic DNA analysis.

II. Related Works

A. Decision Support Systems

Decision Support System (DSS) is earlier defined as a system with a purpose to support management decision-makers in semi-structured decision conditions [11]. It was implied even though not detailed in its early definition that the system must be computer-based, operate online and generate graphical output. DSS has been widely applied for computerized information to ease decision-making in many fields [13]. Hence nowadays, DSS concepts are implemented in multiple domains including forensic [13-16], medicine, military, business and engineering.

B. Visualization and its Applications

FreeViz [2] is applied as a result of concepts combined from statistics, visualization and machine learning because it has some similar concepts to multiple techniques from these fields, while this research project is like a combination between statistics, visualization and Bayesian networks. Demsar et al. [2] presents a new method for intelligent visualization of explorative data analysis in biomedical data.

Aigner et al. [1] set a specific goal to develop visualization and interaction methods for supporting medical personnel in computerized protocol-based care. They considered several data issues that are logic and structures as well, besides patient data in form of parameters and variables. Aigner et al. [1] claim their form of representation is simple, clear and graspable, which become the foundation of visualizations for the representation of data. “The interactive integration of different visualization methods forms a novel way of combining, relating, and analyzing different kinds of medical data.” [1]. Table I summaries several related

TABLE I.
RELATED WORKS TO THIS RESEARCH

Field	Articles	
	Title	Main points
Visualization	CareVis: Integrated visualization of computerized protocols and temporal patient data (Aigner and Miksch, 2006) [1]	Currently, visualization support for patient data analysis is mostly limited to the representation of directly measured data, but explanations is mostly spared out in analysis process – this work aims to fill this gap via integrating classical data visualization and visualization of treatment information.
Visualization	FreeViz: An intelligent multivariate visualization approach to explorative analysis of biomedical data (Demsar et al. 2007) [2]	Visualization can largely improve biomedical data analysis. In a single graph, the resulting FreeViz visualization can provide a global view of the classification problem being studied, reveal interesting relations between classes and features, uncover feature interactions, and provide information about intra-class similarities.
Probability, Statistics and Bayesian networks	Probabilistic graphical models in artificial intelligence (Larranaga, P. and S. Moral, 2011) [10]	Researchers discussed the suitability of probability for intelligent systems, including the role of probabilistic graphical diagrams. Early works in this field and its relevancy is briefly explained as well.
Probability, Statistics and Bayesian networks	A graphical model for the evaluation of cross-transfer evidence in DNA profiles (Colin Aitken, Franco Taroni, Paolo Garbolino, 2003) [12]	Researchers review fundamental issues on the use of graphical models for evidence evaluation. Researchers described mostly with particular reference to the role of cross-transfer evidence. Discussion also involves the determination of factors as nodes, association as links, and probabilities to be included.

works to this research regarding visualization and Bayesian networks.

C. *Bayesian Networks for Forensic DNA Analysis*

Bayesian networks have been applied widely in many practical fields such as medical diagnosis, clinical decision support system, crime risk factors analysis, inference problems in forensic science, and terrorism risk management [17] until present. Bayesian networks can show a very clear relationship of problems in a simple way for statistical analysis.

Mathematical model used for calculation of allele frequencies [18] follow the concept of Bayesian networks [19] while the mathematical model used for calculation of population statistics follow the original concept of Bayes' theorem [19, 20]. For ease of statistical analysis as a result, concept of visualization hence is applied.

III. FRAMEWORK

Forensic DNA analysis includes complex DNA statistics and DNA match probability in criminal cases [9, 21, 22]. In relation to forensic DNA analysis in crime investigation, crime forecasting [23] and knowledge base inference [15] can also be applied. Another researchers that are working together for this project are as on title 'Representing Knowledge for Forensic Evidence Analysis' and 'Implementation of ARIMA model with Fuzzy for Crime Forecasting' [15, 23].

The framework of Crime Management System that we are working on is shown in Figure. 1. The elements that our team concentrates right now are Forensic DNA Analysis, Crime Forecasting and Knowledge Base Inference, while this paper focused on Forensic DNA Analysis as mentioned in Figure 1 and Figure 2.

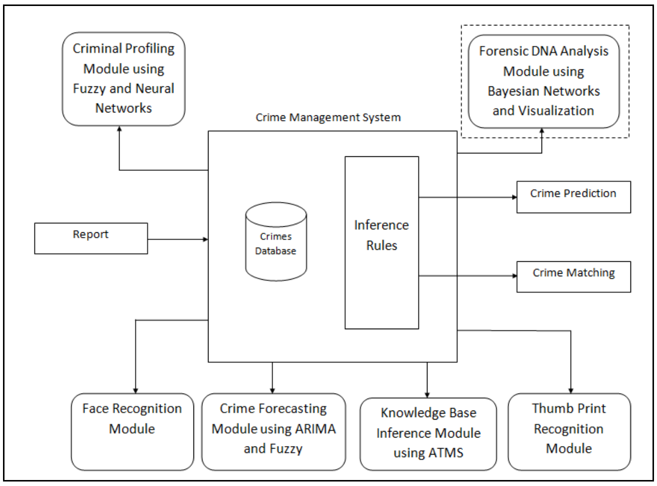

Figure 1. Framework of Crime Management System (CeMaS).

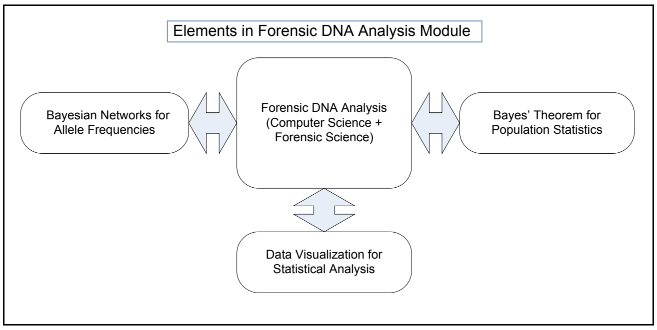

Figure 2. Forensic DNA Analysis (ForAS) Module and its elements.

IV. IMPLEMENTATION

Crime Management System's first web prototype is presented in this section. Screenshot in Figure 3 shows the main user interface for Crime Management System. Less word and more pictures are used to provide usability to forensic scientists and law enforcement.

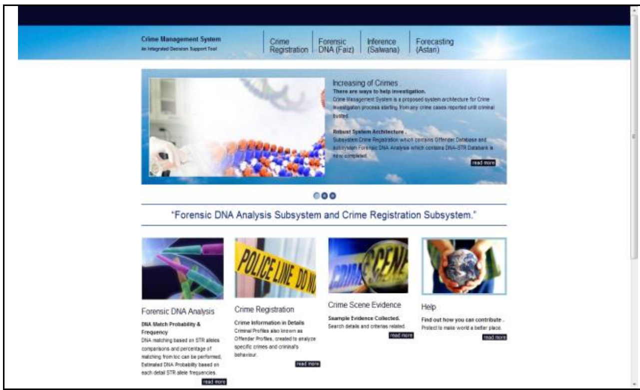

Figure 3. Main page of Crime Management System (CeMaS)

The information in Crime Registration Form is filled up first when crime occurs. This will act as a main entry for Crime Management System. Domain expert (forensic scientists and law enforcements) can also view information about existing crimes that already registered. Figure 4. shows an introduction to Forensic DNA Analysis Module created to brief about the functionality of this module.

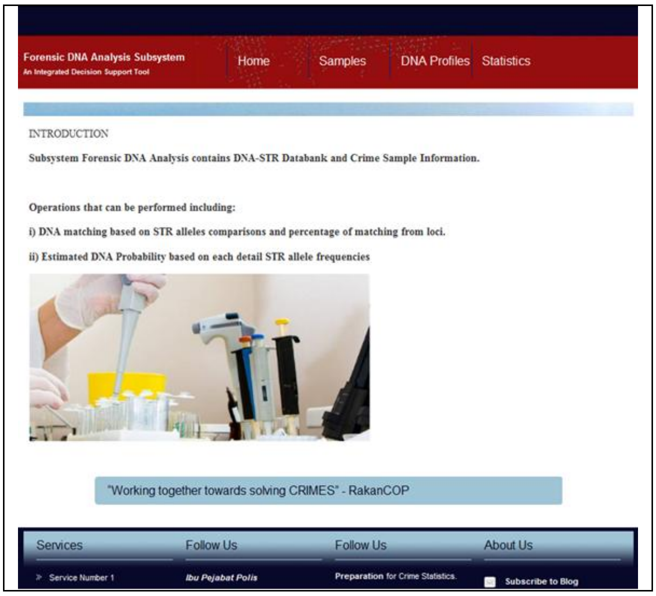

Figure 4. Main page of Forensic DNA Analysis Subsystem (ForAS)

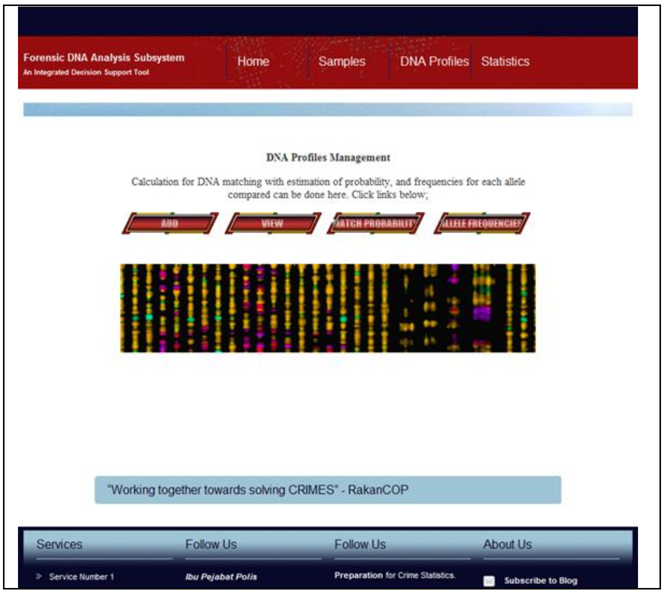

Figure 5. Menu options to handle DNA profiles and DNA analysis

Information about samples collected can be added into database through menu "Samples". The required text field information arranged is based on our study and visit to Forensic DNA Lab of Royal Malaysia Police (RMP). Under the menu "DNA Profiles", analysis can be done once users click the button "MATCH PROBABILITY" or "ALLELE FREQUENCIES" to perform analysis for forensic DNA profiles as shown in Figure 5. Besides, basic functionality such as "ADD" and "VIEW" are also provided. Connection to Google Map API in concept spatial visualization is applied too. Screenshot in Figure 6. shows the first look for this function.

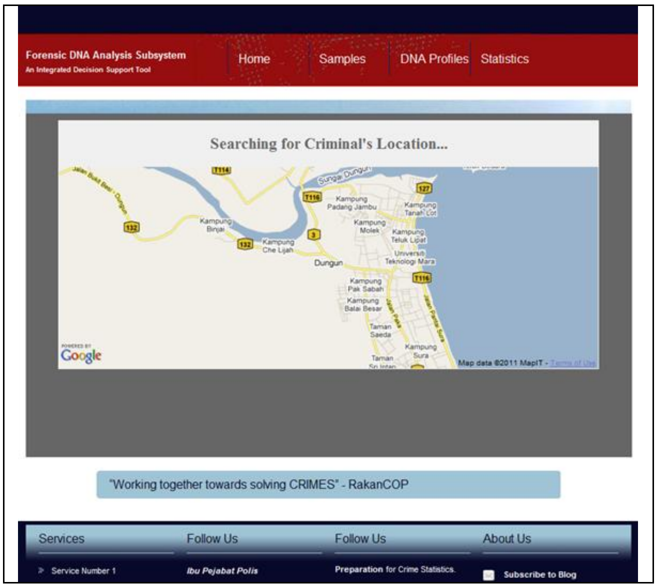

Figure 6. Location-based connection to Google map

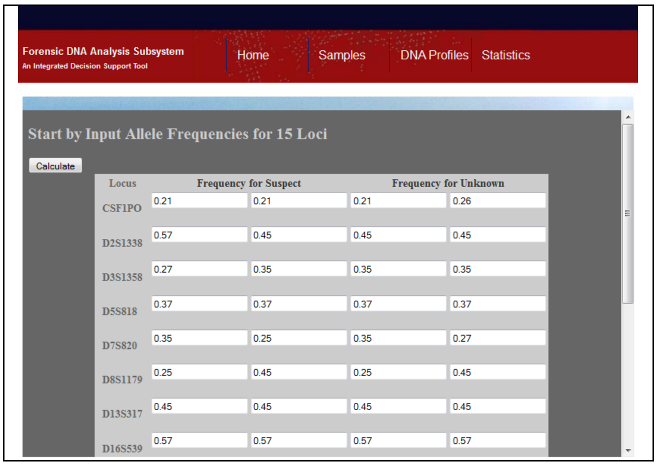

Figure 7. Main form to input DNA allele frequencies for analysis.

Allele frequencies for 15 loci of suspect DNA profile and unknown DNA profile can be entered as shown in Figure 7. to begin the analysis of both DNA profiles. Once all the input submitted, analysis of results will be displayed immediately. Output are including loci frequencies from each alleles, matching for each alleles, estimated DNA profile probability and match rate between DNA profiles of suspect and unknown. Matching for each alleles of comparison will also be marked by green color of table field (to indicate "MATCH") and red color of table field (to indicate "NOT MATCH") as in Figure 8.

Additionally, based on our research and literature reviews on available forensic DNA analysis tools and current visualization tools, there are a strong need for visualization-analysis design, hence here in this project, we show the first look of our visualization-analysis for this tool. When domain experts scroll mouse over each pieces, details about frequencies value will come out as in Figure 9.

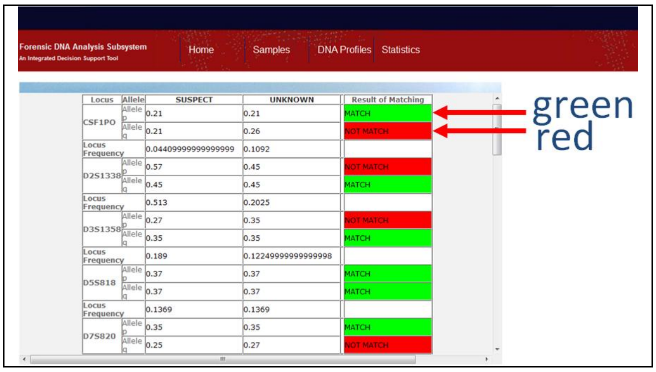

Figure 8. Result of DNA allele frequencies from analysis.

Figure 9. Visualization of allele frequencies from DNA profiles.

Sample search can be done according to type of crimes. Just input the type of crime, and then click Search button. Crime search can be done according to type of crimes, or even just by suspect's name. Just input the type of crime, and then click Search button. Data report with details including crime, suspect, location and time will be displayed. Domain expert can then view all samples information from existing cases. DNA profiles can be chosen based on barCode as the primary key, therefore DNA matching can be performed. In the Profile Entry page, domain expert need to key in values of allele frequencies.

Under DNA profile menus, calculation of DNA matching with estimation of probability and comparison for each allele frequencies can be determined. For a quick summary result (to skip details), domain expert can get direct match rate result instantly through menu MATCH PROBABILITY.

To transform the statistical analysis an immediate visual result, domain expert can start by inputting allele frequencies for suspect and unknown DNA profiles, and then click the submit button. Detail comparisons of allele frequencies, loci frequencies and estimated DNA profile frequencies including match rate will come out for further visualization analysis. Besides DNA data, information of other samples can also be saved. These are including sample type, ethnicity, purpose, gender, profile kit, date of birth, and report number, where all of this information can be searched through a primary key called barCode.

Testing has been done using mock data due to security issue as mentioned by forensic scientists and Royal Malaysian Police (RMP).

One of the benefit of this research is it enables effective data visualization by arranging allele frequencies of forensic DNA analysis results from statistical table into a statistical visualization-analysis in an informative spiral shape diagram like a pie chart. It is also expected to reduce the time taken for evidence evaluation and decision-making (Figure 9).

This kind of visualization is a better way of representation, compared to classical reports and statistical tables. The details of DNA profiles for unknown and suspect can be visualized in just one diagram for the ease of analysis.

Crime Management System web prototype is implemented to be a self-guided forensic DNA analysis system with usability graphical user interfaces rather than classical way that requires training manual and knowledge expertise. The views of results are applied with visualization techniques that are designed to facilitate task handling. Visualization is implemented to represent results while human's interpretation is utilized for driving analysis steps and processes.

On the technical issues, this web prototype is built using Java, an Object-Oriented language, and Java Server Pages web programming language. Java Server Pages can then be implemented together with JavaScript in order to improve interactivity and usability. Aigner et al. [1] also built their prototype using Java as a proof of concept and in order to generate better impression of interactions. Java has many advantages including predefined methods that can be used and adjusted in variety ways dynamically.

CONCLUSIONS

Visualization results in the form of graphical diagram are depicted in such a way so that it is easy to interpret by forensic scientists and law enforcements. Statistical graph and score table is calculated based on a mathematical model that employed the concept of Bayesian networks. This research is significant in supporting decision-making where the concept of multi-disciplinary research field is applied for example between Computer Science and Forensic Science.

The web prototype is implemented to provide assistance in forensic DNA analysis. Bayesian networks and statistical analysis that being employed will reduce time taken for DNA evidence to be analyzed in crime investigation; and plus with visualization, it is expected to ease the complex decision-making through the availability of information into immediate visual result. The web prototype is expected to provide statistical analysis of DNA profiles and automatically analyze into visual results. The implementation hopefully can increase efficiency of crime investigation and its accessibility of information using visualization.

FUTURE WORK

Future work will cover on Criminal Profiling using Fuzzy and Neural Network in Crime Investigation first, before going further to image processing for Face Recognition and Thumb Print Recognition Module.

ACKNOWLEDGMENT

The authors wish to thank Royal Malaysian Police (RMP) especially Dato' Seri Mohd Bakri Zinin, Director of Criminal Investigation Department, Royal Malaysian Police (RMP). This work was supported in part by a grant from Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education (MOHE), vot num. 59201.

REFERENCES

[1] Aigner, W. and S. Miksch, *CareVis: Integrated visualization of computerized protocols and temporal patient data.* Artificial Intelligence in Medicine, 2006. 37(3): p. 203-218.

[2] Demsar, J., G. Leban, and B. Zupan, *FreeViz--An intelligent multivariate visualization approach to explorative analysis of biomedical data.* Journal of Biomedical Informatics, 2007. 40(6): p. 661-671.

[3] Chiu, C.-Y. and A.D. Russell, *Design of a construction management data visualization environment: A top-down approach.* Automation in Construction, 2010. 20(4): p. 399-417.

[4] Ghazali, A.F., et al. *Mapping of Data Visualization into Bayesian Framework: A Case Study of Forensic DNA Analysis*. in *10th International Annual Symposium (UMTAS 2011)*. 2011. Permai Inn Hotel.

[5] Kelleher, C. and T. Wagener, *Ten guidelines for effective data visualization in scientific publications.* Environmental Modelling & Software, 2011. 26(6): p. 822-827.

[6] Kühl, T., et al., *Can differences in learning strategies explain the benefits of learning from static and dynamic visualizations?* Computers & Education, 2011. 56(1): p. 176-187.

[7] Raj, A., C. Hess, and P. Mukherjee, *Spatial HARDI: Improved visualization of complex white matter architecture with Bayesian spatial regularization.* NeuroImage, 2011. 54(1): p. 396-409.

[8] Silva, S., B. Sousa Santos, and J. Madeira, *Using color in visualization: A survey.* Computers & Graphics, 2011. 35(2): p. 320-333.

[9] Drábek, J., *Validation of software for calculating the likelihood ratio for parentage and kinship.* Forensic Science International: Genetics, 2009. 3(2): p. 112-118.

[10] Larrañaga, P. and S. Moral, *Probabilistic graphical models in artificial intelligence.* Applied Soft Computing, 2011. 11(2): p. 1511-1528.

[11] Turban, E., J.E. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*. 2005: Pearson Prentice Hall.

[12] Aitken, C., F. Taroni, and P. Garbolino, *A graphical model for the evaluation of cross-transfer evidence in DNA profiles.* Theoretical Population Biology, 2003. 63(3): p. 179-190.

[13] Noor, N.M.M., et al., *An Integrated DNA Decision Support System Using Grid Environment for Crime Investigation*, in *4th International Symposium on Information Technology 2010 (ITSim'10)*. 2010, Institute of Electrical and Electronics Engineers Inc.: KLCC. p. 636-641.

[14] Ghazali, A.F., et al. *An Integrated Decision Support System for Forensic DNA Analysis on Grid Computing*. in *9th International Annual Symposium on Sustainability Science and Management (UMTAS 2010)*. 2010. Permai Inn Hotel.

[15] Asmara, S.M., N.M.M. Noor, and M.Y.M. Saman. *Representing Knowledge for Forensic Evidence Analysis*. in *10th International Annual Symposium (UMTAS 2011)*. 2011.

[16] Mohamad, S. and N.M.M. Noor, *Decision Making of Forensic Child Sexual Abuse Examination in Intelligent Approach: An Overview*, in *1 st National Doctoral Seminar in Artificial Intelligence Technology (CAIT'2010)* 2010: Universiti Kebangsaan Malaysia.

[17] Pourret, O., P. Naim, and B. Marcot, eds. *Bayesian Networks: A Practical Guide to Applications*. STATISTICS IN PRACTICE, ed. L. John Wiley and Sons. 2008, WILEY: University of Glasgow, United Kingdom.

[18] Noor, N.M.M. and R. Baital, *Integrated DNA Support System for Crime Investigations (DNA TRACERS).* Journal of Ultra Scientist of Physical Sciences, 2009. 21(3): p. 989-998.

[19] Biedermann, A. and F. Taroni, *Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature.* Forensic Science International: Genetics, 2011. In Press, Corrected Proof.

[20] Dawid, A.P., J. Mortera, and P. Vicard, *Representing and solving complex DNA identification cases using Bayesian networks.* International Congress Series, 2006. 1288: p. 484-491.

[21] Lee, J.W., et al., *Evaluation of DNA match probability in criminal case.* Forensic Science International, 2001. 116(2-3): p. 139-148.

[22] Henderson, J.P., *The use of DNA statistics in criminal trials.* Forensic Science International, 2002. 128(3): p. 183-186.

[23] Astari, et al. *Implementation of ARIMA Model with Fuzzy for Crime Forecasting*. in *10th International Annual Symposium (UMTAS 2011)*. 2011.

Noor Maizura Binti Mohamad Noor was born in Malaysia. She obtained her Diploma and Bachelor of Computer Science from Universiti Pertanian Malaysia, Serdang Selangor in 1991 and 1994 respectively. She received her Master of Science (Computer Science) from Universiti Putra Malaysia in 1997. She later acquired her doctoral degree in Computer Science from University of Manchester, United Kingdom in 2005. Her recent research work focuses on improving organizational decision making practices through the use of technologies. She is a member of Malaysian Software Engineering Interest Group (MySIG), Association for Computing Machinery (ACM) and International Association of Computer Science and Information Technology (IACSIT). Her research interests are 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 started her career as a tutor in 1994 at the Department of Computer Science, Faculty of Science and Environmental Studies, Universiti Pertanian Malaysia (UPM), Serdang, Selangor. Three (3) years later she was appointed as a lecturer before being transferred to Universiti Putra Malaysia Terengganu (UPMT), Kuala Terengganu in November 1997. In 2010, upon her excellent achievements, she has been appointed as an Associate Professor. Currently, she is a deputy director of Quality Centre and Academic Development (PKPA) for Universiti Malaysia Terengganu (UMT). She has been appointed as the head of Computer Science Department for three (3) years (April 2007 – April 2010).

She has extensively collaborated with research institutions and universities in Malaysia. She has received research grants to the tune of more than RM One Million from various agencies either as main or co-researchers, out of which, about RM 346, thousands as project leaders. Her works have been funded by Universiti Malaysia Terengganu (UMT), Ministry of Higher Education (MOHE) and Ministry of Science, Technology and Innovation (MOSTI), Malaysia. She has presented and published a number of papers on the decision support system at various international and local refereed journals, conferences, seminars and symposiumsHer contribution to scientific computing has been recognized by receiving several awards including the Best Scientist Award (2009), Product and Research Innovation Award for International and National Level (2008, 2009 and 2010), UMT Excellence Service Award (2008, 2009, 2010) and Excellent Teaching Award (2006).