Machine Learning: A Means of Diagnosing and Prescribing Treatments for Tuberculosis Patients

Arman James E. Avenir, Jonas D. Villota, and Maryli F. Rosas, *Member, IACSIT*

***Abstract*—Tuberculosis (TB) remains one of the leading health problems in the Philippines. Although curable, many Filipinos cannot afford the cost of treatment. Furthermore, the free services offered by the public health centers are insufficient to attend to the medical needs of those seeking help. In this paper, the researchers present the system that can assist doctors, nurses and health workers in diagnosing tuberculosis using techniques in machine learning through applying ID3 algorithm. Interviews with several experts in the field of TB were conducted in order to gather data that were used to populate the system's knowledge base. Test result show that the system is capable of prescribing treatments based on the patient's data and tracking the progress of the patient based on his/her prescribed treatment.**

***Index Terms*—Diagnostic, machine learning, tuberculosis, treatment.**

I. Introduction

According to the World Health Organization, in 2012, "the largest number of new TB cases occurred in Asia, accounting for 60% of new cases globally [1], with that the Philippines belong to the countries with high Tuberculosis cases. As of April 2012, Ms. Nadal cited in his article that the Philippines carry the 9th largest burden of Tuberculosis in the world. She further stated that in the 2010 data it showed that approximately 390,000 Filipinos have Tuberculosis, and 75 people die daily from this disease [2]. "Tuberculosis (TB) is a common lethal, infectious disease caused by various strains of bacteria, usually Mycobacterium tuberculosis" defined by Tim Sandle [3]. It typically attacks the lungs, which later on affect other parts of the body. It spreads through the air when people with active TB infection cough, sneeze, or otherwise transmit their saliva through the air [4]. Most infections are asymptomatic and latent, only one, out of ten latent infections eventually progresses to active disease which, if left untreated, kills more than 50% of those infected [5].

In Bacoor, Cavite, the case of TB is increasing. In 2010, there were 404 TB cases recorded. The following year, there were 589 TB cases recorded. In 2012, according to Dr. San Luis of the Health Center in Panapaan 1 Bacoor, Cavite, by the end of the year there would be an estimated of 600 plus cases of TB in Bacoor. TB is curable, however not all of the patients can afford the medicine for it. Because of that, people go to the nearest public health center for open service. However, based on the observations and interviews conducted by the researchers in Bacoor Cavite, the number of health professionals does not suffice the need of patients. A doctor cannot also focus with one patient since he/she needs to attend to the massive number of patients registered during that day. Further, when the doctor is not around, the nurse and health workers take over in diagnosing and treating the patients by referring to the doctor's module which consumes more time. The accuracy of diagnosis, generating and administering treatment schedule of regimen are also some of the problems encountered.

In this paper, the researchers explored the use of machine learning techniques to provide support to doctors, nurses and health workers in diagnosing and prescribing treatment to people with TB. machine learning is a branch of Artificial Intelligence that focuses on algorithms and methods to allow a computer to learn through examples by capturing characteristics of interest based on underlying probability distribution [6].

Machine learning has the same concept as Expert system; it can imitate the capabilities of a human expert. It can make an automated decision based on the knowledgebase inputted by the domain expert. Because human expertise is not always available or sufficient to address the needs of the community, a diagnostic software using machine learning can be used as a substitute to the human expertise. As stated on the work of Co [7], that this expert system serves as solutions or repeatedly occurring problems. Further, the study of Bautista-Navarro [8] and group motivates the researchers in using 2 major components of an expert system: knowledgebase database and inference engine which contain rules in making decisions. In addition, the study named THESSA [9] uses Database Management System to store and track records in which the researcher also take full advantage of. Just like the researcher's study, it also uses Expert system in processing schedules and selecting appropriate persons.

II. System Overview

Fig. 1 shows the architectural design of the system. Diagnosing a TB case proceeds as follows. First, patient information consisting of the basic details (name, address, age, height, weight) are fed to the system. Signs and symptoms are also provided, including the sputum examination, the classification of TB, and the patient type. These inputs are needed in order to determine the type of

Manuscript received February 20, 2014; revised May 17, 2014. This work was supported in part by the Faculty Development Fund of the De La Salle University – Dasmarinas under the College of Science and Computer Studies.

Arman James E. Avenir is with Aegies People Support Center, Ayala Ave Cor. Gil Puyat Ave Makati City, Philippines (e-mail: armanavenir@gmail.com).

Jonas D. Villota is with Blinkup Inc. 1404 Prime Land Tower, 2218 Market Street Madrigal Business Park Ayala Alabang, Muntinlupa City, Philippines (e-mail: jonsader.14@yahoo.com).

Maryli F. Rosas is with the Computer Studies Department of De La Salle University – Dasmarinas, City of Dasmarinas Cavite, Philippines 4115 (e-mail: ilyramrosas@yahoo.com).

treatment regimen that will be given to the patient.

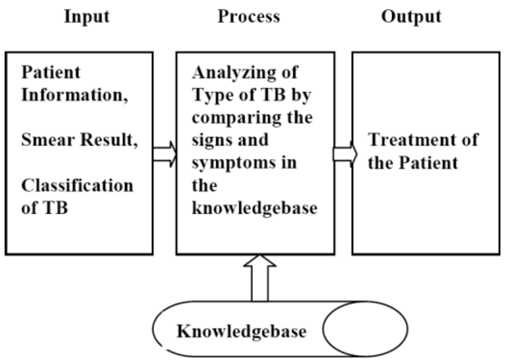

Fig. 1. Architectural design.

The system determines the possible types of TB by comparing the inputs with the data stored in the knowledge base. Once the type has been identified, the system suggests a possible treatment, which include the medicine and the schedule to intake the medicine.

This study aims to assist and unburden the doctors, nurses, and staffs of medical center. Part of the study is the creation of machine learning and how artificial intelligence is applied in diagnosing and treating Tuberculosis. "Artificial intelligence (AI) is a branch of computer science concerned with making computers behave like humans" [10].

The study focused on machine learning that gain knowledge through supervised learning. Supervised learning, imply that an expert is present to define correct and incorrect responses [11]. The machine learning in this study focuses on automated decision, based on recognized information in the knowledgebase of the system. The information gathered in this study will be limited only in Bacoor, Cavite particularly in the Health Center of Panapaan 1 Bacoor, Cavite. Furthermore, the system can only be used by the doctors and staffs of the health center.

This study covered the aspects of TB based on the doctor's module. Currently, there are two (2) classification of TB namely: Pulmonary, which happens in the lungs and Extra pulmonary, which happens in other parts of the body with the exception of lungs. In order to cure these types of TB, doctors prescribe a treatment regimen to the patient. A treatment regimen is a set of schedule and medicine for the patient to follow. There are 3 types of treatment regimen, Category 1 regimen which is prescribed to new patients with positive smear findings or to new patients with extra pulmonary TB. Category 2 regimen, which is prescribe to previously treated patients. And Category 3 regimen, which is prescribe to new patients with negative smear findings, pulmonary TB, and with minimal parenchymal lesions on chest x-ray.

The specialists interviewed in this study are the doctors of Health Centers with Direct Observe Treatment Short Course (DOTS) facility in Bacoor, Cavite. A health center with DOTS facility is specially designed for treating patients with TB. They are Doctor San Luis the City Health Officer, Doctor Michael Angelo Marques of DOTS 1 Facility, and Doctor Grace Aceron of DOTS 2 Facility.

III. Concept of the Study

This study uses the concept of supervised learning, decision trees, database for knowledge-based learning, and ID3 Algorithm.

The need to provide faster and easier access to users and managing complex data are objectives of database [12] that the researcher took advantage of in developing the system for the this study.

Another concept used in the study is the supervised learning which implies that an expert is present to input a set of training data. The training data contains both a predictor (independent variable) and target (dependent variable) whose value is to be estimated. Through the process of supervised learning, it can predict the value of the target variable based on the predictor variables [12].

Decision Trees is one of the most intuitive and practical methods for supervised learning. A decision tree is the result of a classification process in which the source data is reduced into a predictor tree that represents a set of if/then/else rules. The advantages of decision trees is that they are very fast (tend to classify in fewer decisions than the features of the data set) and are also easy to interpret.

Further, the core algorithm for building decision trees called **Iterative Dichotomiser 3 (ID3) Algorithm** by J. R. Quinlan which employs a top-down, greedy search through the space of possible branches with no backtracking. ID3 uses Entropy and Information Gain to construct a decision tree [13], [14].

The ID3 algorithm can be summarized as follows:

- Take all unused attributes and count their entropy concerning test samples.
- Choose attribute for which entropy is minimum (or, equivalently, information gain is maximum)
- Make node containing that attribute.

Entropy is used to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. Further, Entropy is used to determine which node to split next in the algorithm. The higher the entropy, the higher the potential to improve the classification.

The formula for computing the Entropy using the frequency table of one attribute is shown in (1):

$$E(S) = p_1 \log 2 p_1 - p_2 \log 2 p_2 - \ldots \, p_n \log 2 p_n \quad (1)$$

where: p set of samples (e.g. p_1 = no. of yes/ total no. of set.
p_2 = no. of no/total no. of set)
S is the sample of attribution.

While, Entropy using the frequency table of two attributes may be computed as seen in (2):

$$\text{Entropy}\,(T, X) = \sum c \in \chi P(c) E(c) \quad (2)$$

where: $P(c)$ = average of the sample.
$E(c)$ = entropy of the sample.

The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain. The attribute with the largest information gain is used as the decision node of the decision tree.

Equation (3) shows formula for Information Gain:

$$\text{Gain}\,(T, X) = \text{Entropy}\,(T) - \text{Entropy}\,(T, X) \quad (3)$$

The following information is used as a test data in the proposed system's ID3 algorithm:

The gathered test data are from the doctor's TB module provided by Doctor San Luis III. By using these test data in the algorithm, the ID3 recognizes similar patterns and generates the following conditions:

R1: If (Type of Patient = Others) Then
Treatment = Category 2
R2: If (Type of Patient = Relapse) Then
Treatment = Category 2
R3: If (Type of Patient = Return After Default) Then Treatment = Category 2
R4: If (Type of Patient = Treatment Failure) Then Treatment = Category 2
R5: If (Type of Patient = New) AND
(Type of Smear = Smear Positive) Then
Treatment = Category 1
R6: If (Type of Patient = New) AND
(Type of Smear = Smear Negative Extensive) Then Treatment = Category 1
R7: If (Type of Patient = New) AND
(Type of Smear = Smear Negative Minimal) AND (Type of TB = Pulmonary) Then
Treatment = Category 3
R8: If (Type of Patient = New) AND
(Type of Smear = Smear Negative Minimal) AND (Type of TB = Extra Pulmonary) Then
Treatment = Category 1

From the conditions generated from the ID3 algorithm, the resulting decision tree is shown in Fig. 2.

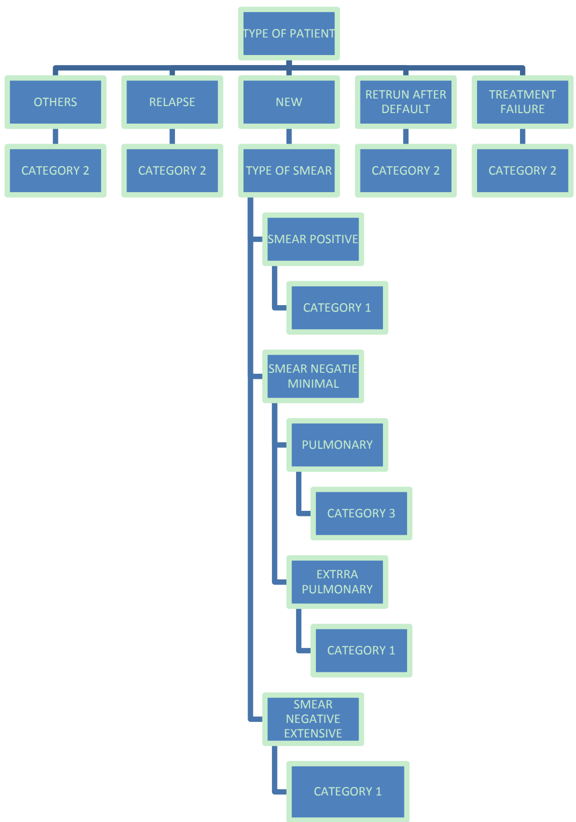

Fig. 2. Resulting decision tree.

While Table I summarizes the type of patient, the possible diagnosis and treatment category which serves as the test data for this study.

TABLE I: TABLE OF TEST DATA

Type of Patient	Type of TB	Type of Smear	Treatment
NEW	PULMONARY	SMEAR POSITIVE	CATEGORY 1
NEW	PULMONARY	SMEAR NEGATIVE EXTENSIVE	CATEGORY 1
NEW	EXTRA PULMONARY	SMEAR NEGATIVE MINIMAL	CATEGORY 1
NEW	EXTRA PULMONARY	SMEAR POSITIVE	CATEGORY 1
NEW	EXTRA PULMONARY	SMEAR NEGATIVE EXTENSIVE	CATEGORY 1
NEW	PULMONARY	SMEAR NEGATIVE MINIMAL	CATEGORY 3
RELAPSE	PULMONARY	SMEAR POSITIVE	CATEGORY 2
TREATMENT FAILURE	EXTRA PULMONARY	SMEAR NEGATIVE MINIMAL	CATEGORY 2
OTHERS	EXTRA PULMONARY	SMEAR NEGATIVE EXTENSIVE	CATEGORY 2
RETURN AFTER DEFAULT	EXTRA PULMONARY	SMEAR POSITIVE	CATEGORY 2

To compute for the entropy of the variable Treatment, (4) will be used.

$$\text{Entropy}(S) = \sum -P_i \text{Log2 } P_i \tag{4}$$

where: p set of samples,
p_1 = no. of records for each type/ total no. of records.
p_2 = no. of no/total no. of set.
S is the sample of attribution.

Based on Table I, treatment falls on three categories. Further, there are five occurrences of Category 1, four of Category 2 and one of Category 3. Hence, the formula may be expounded to (5). Substituting these values to the variable Pi the following will be computed (6), (7), (8), (9):

$$\text{Entropy(Treatment)} = \text{Entropy}(5,4,1) \tag{5}$$

$$\text{Entropy(Treatment)} = -(5/10\text{Log2}(5/10)) - (4/10\log2(4/10)) - (1/10\log2(1/10)) \tag{6}$$

$$\text{Entropy(Treatment)} = -(0.5\log2(0.5)) - (0.4\log2(0.4)) - (0.1\log2(0.1)) \tag{7}$$

$$\text{Entropy(Treatment)} = 0.5 + 0.53 + 0.33 \tag{8}$$

$$\text{Entropy(Treatment)} = 1.36 \tag{9}$$

Therefore, the Entropy for Treatment is 1.36 shown in (9).

Test data comprises of Types of Patient, Type of TB and Type of Smear. The following equation (10) (11) (12) are the equation for the Entropy in the test data:

$$\text{Type of Patient} = \text{Entropy(Treatment,Type of Patient)} \quad (10)$$

$$\text{Type of TB} = \text{Entropy(Treatment,Type of TB)} \quad (11)$$

$$\text{Type of Smear} = \text{Entropy(Treatment, Smear)} \quad (12)$$

To get the Entropy of each type, we will follow the (13)

$$\text{Entropy}(T,X) = \sum P(c)E(c) \quad (13)$$

where: T = Treatment ; X = Type of Patientor X = Type of TB or X = Type of Smear

To compute for the Entropy of Treatment for each Type of Patient, we first have to compute for the entropy of each Patient and we will follow the (4)

$$\text{Entropy (New)} = \text{Entropy}(5,0,1) = (-[(5/6)*\log2(5/6)] + -[(0/6)*\log2(0/6)]) + -[(1/6)*\log2(1/6)] \quad (14)$$

There are five occurrences of New Patient under Category 1, zero instances under Category 2 and 1 occurrence under Category 3. Substituting the values shown in (14) will give us 0.39 value. While for other types of patients such as Others, Relapse, Return after Default and Treatment Failure has zero instances of Category 1, one instance of Category 2, and zero instance of Category 3 which yields to 0.00 result.

Table II summarizes the Entropy and Result of all Type of Patients.

The Entropy for the Treatment of different type of Patient is shown in (15), (16) and (17)

TABLE II: TYPE OF PATIENT

Patient Type	**Entropy**	**Result**
New	5,0,1	0.39
Others	0,1,0	0.00
Relapse	0,1,0	0.00
Return After Default	0,1,0	0.00
Treatment Failure	0,1,0	0.00

$$\text{Entropy(Treatment, Type of Patient)} = P(\text{New})*\text{Entropy}(5,0,1) + P(\text{Others})*\text{Entropy}(0,1,0) + P(\text{Relapse})*\text{Entropy}(0,1,0) + P(\text{Return After Default}) * \text{Entropy}(0,1,0) + P(\text{Treatment Failure}) * \text{Entropy}(0,1,0) \quad (15)$$

$$\text{Entropy(Treatment,Type of Patient)} = 0.39 + 0 + 0 + 0 + 0 \quad (16)$$

$$\text{Entropy(Treatment,Type of Patient)} = 0.39 \quad (17)$$

Following the same formula (4) for each variable X, (now Type of TB), the following will be the result (shown in Table III).

TABLE III: TYPE OF TB

Type of TB	**Entropy**	**Result**
Pulmonary	2,1,1	0.6
Extra Pulmonary	3,3,0	0.6

The Entropy for the Treatment of different type of TB is shown in (18) (19) and (20).

$$\text{Entropy (Treatment, Type of TB)} = P(\text{Pulmonary}) *\text{Entropy}(2,1,1) + P(\text{Extra Pulmonary})*\text{Entropy}(3,3,0) \quad (18)$$

$$\text{Entropy (Treatment, Type of TB)} = 0.6 + 0.6 \quad (19)$$

$$\text{Entropy (Treatment, Type of TB)} = 1.2 \quad (20)$$

While for the Type of Smear, the same formula (4) will be used. Table IV shows the summary of result.

TABLE IV: TYPE OF SMEAR

Type of Smear	**Entropy**	**Result**
Smear Positive	2,2,0	0.40
Smear Negative Extensive	2,2,0	0.28
Smear Negative Minimal	1,1,1	0.48

The Entropy for the Treatment of different type of Smear is shown in (21), (22), (23) and (24)

$$\text{Entropy } (T, X) = \sum P(c)\, E(c)$$

where: T = Treatment; X = Type of Smear

$$\text{Entropy(Treatment,Type of Smear)} = P(\text{Smear Positive})* \text{Entropy}(2,2,0) + P(\text{Smear Negative Extensive})* \text{Entropy}(2,1,0) + P(\text{Smear Negative Minimal})* \text{Entropy}(1,1,1) \quad (21)$$

$$\text{Entropy(Treatment,Type of Smear)} = P(\text{Smear Positive})* \text{Entropy}(2,2,0) = (4/10) * (-[(2/4)*\log2(2/4)] + -[(2/4)*\log2(2/4)] + - [(0/4)*\log2(0/4)]) \quad (22)$$

$$\text{Entropy (Treatment, Type of Smear)} = 0.4 + 0.28 + 0.48 \quad (23)$$

$$\text{Entropy (Treatment, Type of Smear)} = 1.16 \quad (24)$$

Moving forward, getting the Information gain of each type of Entropy, equation (25)

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$
$$\text{Gain(Treatment, Type of Patient)} = \text{Entropy (Treatment)} - \text{Entropy(Treatment, Type of Patient)} \quad (25)$$

$$\text{Gain(Treatment, Type of Patient)} = 1.36-0.39 \quad (26)$$

$$\text{Gain(Treatment, Type of Patient)} = 0.97 \quad (27)$$

Table V shows the summary of the computed Information Gain.

TABLE V: INFORMATION GAIN

X	***T*= Treatment**	**Entropy(*X*)**	**Result**
Type of Patient	1.36	0.39	0.97
Type of TB	1.36	1.20	0.16
Type of Smear	1.36	1.16	0.20

Note: The largest information gain will be the base node, if a certain type has same values then the result is the possible same values as leaf node else get the next largest value to be part as root node.

The reason why researchers used the ID3 algorithm is that they will use it to determine what kind of treatment will be given to the patient. It will be based on the type of patient, type of smear and type of tuberculosis. Also by using the ID3 algorithm, there is a proof that the system is learning as it will generate its own decision trees based on the user inputs.

IV. Summary and Findings

To validate the correctness of the system in diagnosing and prescribing treatments, a group of doctors, nurses and midwives from the health center of Panapaan 1 Bacoor, Cavite participated. Survey questions were administered to gather the respondents' feedback regarding the system. There were 10 respondents in all, 3 of which are doctors, 2 are nurses, and 5 are midwives.

The survey is composed of 11 questions, answerable by 5 levels of ratings namely: Outstanding, Very Satisfactory, Satisfactory, Needs Improvement and Poor. The questions are divided into three (3) categories namely: Usability, Reliability and Security. Respondents were asked to evaluate the program's performance based on their experience in using the program. The respondent is required to choose only one among of the 5 ratings. Comments and suggestions were also provided at the bottom of the questionnaire.

Usability section pertains to the system's user interface, presentation of data, generation of reports, ease of use of the system, and efficiency of program. Reliability section contains the system's effectiveness with regard to decision making, consistency of data, and accuracy report. Finally, the Security category covers how well the system protects the patient records, accounts and the knowledgebase of the system.

TABLE VI: Overall Rating per Attribute

Attributes	Overall Rating (Arithmetic Mean)
Usability	4.54
Reliability	4.76
Security	4.63
Total Average	**4.64**

In Terms of Usability, the system received an average score of 4.54. This means that any type of user with different computer learning ability and skills can use the system effectively and efficiently.

The system received an average score of 4.76 in terms of Reliability as a decision making tool. The generated diagnosis and patient treatment schedule is reliable and this can make use full advantage of it to monitor religiously the different types of patients.

While in terms of Security, the system received an average score of 4.63. This means that even though there are several users who can access the system, the user level are properly imposed with different access rights to each modules of the system.

Table VI shows the Overall Rating per Attribute as well as the Total Average Mean of 4.64 was gained.

V. Conclusions

In conclusion, the authors of this study were able to create a machine learning software on diagnosing and treating TB by applying artificial intelligence. The researchers were able to gather information on diagnosing and treating TB by conducting an interview with the experts in the field of TB. With the information they have learned from the TB experts, they are able to apply it in the knowledgebase of the system. Thus, in developing this program, the rapid prototyping [15], [16] approach is applied wherein the researchers must continually consult with the TB expert and check if the system meets the requirements.

Test Result shows that the system got an overall average rating of 4.64 which means that the system is effective and efficient in terms of Usability, Reliability and Security.

Based on the comments received by the researchers from the experts of TB, a software in diagnosing and treating TB can speed up the work of the user in diagnosing and treating TB as it can help them decide faster and unburden their works. However, in order to use this program the user must have at least a basic knowledge in TB.

Finally, more tests must be conducted to ensure program constancy. Preferably, these tests must be conducted in an actual environment that is to all health centers of Bacoor, Cavite. The trial phase should run for at least a month to identify and fix all possible problems of the software.

In conclusion, the authors of this study were able to create a machine learning software on diagnosing and treating TB by applying artificial intelligence. The researchers were able to gather information on diagnosing and treating TB by conducting an interview with the experts in the field of TB. With the information they have learned from the TB experts, they are able to apply it in the knowledgebase of the system. Thus, in developing this program, the rapid prototyping [15], [16] approach is applied wherein the researchers must continually consult with the TB expert and check if the system meets the requirements.

Test Result shows that the system got an overall average rating of 4.64 which means that the system is effective and efficient in terms of Usability, Reliability and Security.

Based on the comments received by the researchers from the experts of TB, a software in diagnosing and treating TB can speed up the work of the user in diagnosing and treating TB as it can help them decide faster and unburden their works. However, in order to use this program the user must have at least a basic knowledge in TB.

Finally, more tests must be conducted to ensure program constancy. Preferably, these tests must be conducted in an actual environment that is to all health centers of Bacoor, Cavite. The trial phase should run for at least a month to identify and fix all possible problems of the software.

Acknowledgment

The authors of this research wish to express their utmost gratitude and appreciation to the following individuals:

To Ms. Sherry B. Naz, Ms. Emelyn D. Mayuga, Ms. Mary Anne C. Paguio, Mr. Paulino H. Gatpandan, Mr. Rolando B. Barrameda, Dr. Johnny A. Ching, Ms. Olivia M. Legaspi and Dr. Willington Onuh for their utmost support.

To Mayor Strike Revilla for giving them an opportunity to help the City Health Office of Bacoor. To Dr. Reynaldo G.

San Luis III and the City Health Office staffs of Bacoor for their unending patience and assistance. Thank you very much!

To their parents, husband and kids for their unending support towards the end of this research. Their patience, love and understanding which encouraged the researchers to finish this research. And also for the financial support they have given us. Thank you!

And most especially, to the heavenly Father for His guidance. Without Him all of this will never be possible.

References

[1] Tuberculosis. (October 2013). *Fact Sheet N104*. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs104/en/

[2] R. Nadal. (2012). RP's TB control program strengthens as USAID reaffirms support to NTP. [Online]. Available: http://www.pchrd.dost.gov.ph/index.php/2012-05-23-07-46-36/2012-05-24-00-01-11/4896-rp-s-tb-control-program-strengthens-as-usaid-reaffirms-support-to-ntp

[3] T. Sandle, "Drug Resistant TB cases are 'Underestimated'," WHO, 2013.

[4] E. Andam, "Analysis of transmission dynamics of tuberculosis (tb) usingdifferential equations: A case study of amansie west district, Ghana," Master Thesis, wame Nkrumah University of Scienceand Technology, 2013.

[5] Tuberculosis. *LenMed Health*. [Online]. Available: http://www.lenmedhealth.co.za/lenart1.htm

[6] J. Burkhow, "What is Machine Learning?" *Data Enthusiast*, August 1, 2011.

[7] C. B. Co, "An expert system for semiconductor device assembly diagnostics," Dissertation, De La Salle University, Manila, 2010.

[8] R. T. Batisa-Navarro, D. A. Bandojo, M. Gatapia, R. N. Santos, A. B. Marcelo, L. C. Panganiban, P. C. Naval, "ESP: An expert system for poisoning diagnosis and management," *Inform Health Soc Care*, March 2010.

[9] A. C. H. Cortez, C. A. M. Nacu, C. R. B. Payos, B. Tate Y. Sy, and R. K. Villanueva, "Thesis system support application (THESSA) with expert system integrated in Joomla," thesis, De La Salle University, Manila, 2010.

[10] Artificial Intelligence. (N.D.). [Online]. Available: http://www.webopedia.com/term/a/artificial_intelligence.html

[11] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed, New Jersey: Prentice Hall Inc., 2009.

[12] J. Hoffer, M. Prescott, and F. McFadden, *Modern Database Management*, 8th ed., Pearson Education, 2008.

[13] Sayadsaed. (2012). An introduction to data mining. [Online]. Available: http://www.Saedsayad.Com/Decision_Tree.Htm

[14] Inference Engine, *Encyclopedia Britannica Online*. [Online]. Available: http://www.Britannica.Com/Ebchecked/Topic/287556/Inference-Engine

[15] Prototyping Model. *Toolbox.Com Wiki Online*. [Online]. Available: http://it.toolbox.com/wiki/index.php/prototyping_model

[16] Software Prototyping. (2012). *Wikipedia Online*. [Online]. Available: http://en.wikipedia.org/wiki/software_prototyping

Arman James Escarez Avenir was born in Las Piñas city, Philippines on April 20, 1992 with the citizenship of Filipino. Avenir studied in De La Salle University – Dasmariñas in 2009 and graduated in 2013. He took the bachelor of science in computer science which focuses mostly in computer theory, analytical thinking and research.

He received his on the job training as a computer technician/troubleshooter in FilinvestAlabang Inc in 2012. His past job experience primarily focused on network management, PC troubleshooting and data warehouse management. As of 2014, he works on aegis people support located in Aegis People Support Center, Ayala avenue corner Senator Gil Puyat Avenue, Makati City, Philippines, 1200 with the position of Specialist, Software Engineering. His current job focuses on business intelligence, reporting services and SQL programming.

Mr. Avenir became a member and officer of Lasallian Computer Society in the De La Salle University -Dasmariñas from 2010 to 2013. He became a member of Junior Programming Computer Society from 2010 to 2011.

Jonas De Guzman Villota was born in Jeddah, Saudi Arabia on February 14, 1993 with the citizenship of Filipino. Villota studied in De La Salle University – Dasmariñas in 2009 and graduated in 2013. He took the bachelor of science in computer science which focuses mostly in computer theory, analytical thinking, and research.

He received his on the job training as a programmer/encoder in the Municipality of Bacoor, Cavite in 2012. His past job experience focuses in data management and programming. As of 2014, he works in Blink Up Inc. located in 1404 Prime Land Tower, 2216 Market Street, Madrigal Business Park, Ayala Alabang, Muntinlupa City with the position of I.T Developer. His current job focuses on programming and map development.

Mr. Villota became a member of Lasallian Computer Society in the De La Salle University – Dasmariñas from 2012 to 2013.

Maryli Feliciano Rosas was born in Paranaque Philippines on September 9, 1976. Rosas receives her bachelor degree in computer science at Philippine Christian University, Philippines in 1997. Rosas also earned her master degree in information technology at Adamson University, Philippines in 2005. Rosas further pursued her doctor degree in information technology at AMA University and currently on dissertation. She is currently connected with the Computer Studies Department, De La Salle University – Dasmarinas, Philippines and works as a full time associate professor. She was also the former Department chair and computer laboratory supervisor from April 2010 up to May 2012 on the same university. She co-authored a completed research project titled: Worktext in Data Stuctures and Algorithm, co-proponent of applied research funded by the University Faculty Research Office entitled: A Computerized Entrance Examination for De La Salle University – Dasmarinas and currently working on a tracer study on I.T and C.S. graduates of DLSU-D. She is also a co-presenter at International Congress on e-Learning 2013 entitled: DB2CAI: An E-Learning Experience on the Web and Smartphones.

Ms. Rosas is a member of IACSIT, PSITE, MRSP, PeLS and CSP-SIGNLP.