Abaca Tissue Culture Contamination Grading Using Naïve Bayesian Classification

Rhoderick D. Malangsa and Elmer A. Maravillas

***Abstract*— One of the challenges meet in the agricultural sector is the plant disease which affected food crops, causing significant losses to farmers and its spread has increased dramatically in recent years. The abaca plantation industry in the province of southern was greatly affected by the disease and was estimated to suffer about 30 per cent in damages. There have been many studies before utilizing Naïve Bayesian algorithm on disease and contamination prediction especially for crops but the study on *in vitro* abaca specimens has not given enough focused yet. This study is created for a contamination grading for abaca tissue culture specimen using Naïve Bayesian classification. In phase one, capturing specimen images thru masking using a camera. Phase two was feature extraction techniques of RGB mean values and whitish binary image to obtain relevant data to be used in phase three and four where specimens were classified as either healthy or contaminated. Lastly, the performance of these classifiers was evaluated based on the overall accuracy, precision, and specificity. The overall accuracy of the system was 90% in contamination grading. Moreover, the values generated in precision, recall, and specificity also indicate the very good performance of the classifier. The study indicated that Naive Bayes has good potential for identifying contamination grading accurately that mainly causes by fungi in abaca tissue culture laboratory.**

***Keywords*—Classification, Contamination Grading, Expert System, Naïve Bayesian.**

I. Introduction

Agriculture is one of the most ancient activities of man in which innovation and technology are usually accepted with difficultly, unless real and immediate solutions are found for specific problems or for improving production and quality. Nevertheless, a new approach, of gathering information from the environment, could represent an important step towards high quality and eco-sustainable agriculture [1].

One of the challenges meet in the agricultural sector is the plant disease. Plant disease affect food crops, causing significant losses to farmers and the spread of plant diseases has increased dramatically in recent years.

On 2003, a certain disease threatens the abaca industry in the province of southern Leyte. Eighty percent of the province's abaca plantation particularly in Sogod town was greatly affected while Maasin City was estimated to suffer about 30 percent in damages [3]. With the presence of diseases in the field, it is difficult to propagate disease-free planting materials using the conventional method.

One of the possible solutions is tissue culture, and is seen as an important technology for developing countries for the production of disease-free, high quality planting material and the rapid production of many uniform plants. In this way, thousands of copies of a plant can be produced in a short time [2]. This project involves the mass production of recommended and disease-free abaca planting materials through tissue culture using the shoot tip technique. The abaca tissue culture laboratory is a controlled sterilized laboratory.

However, manual visual inspection of the in vitro specimen inside the laboratory will also trigger the contaminants in abaca tissue culture laboratory. Frequently encountered bacterial and fungal contamination especially in laboratories of micro propagation posed a considerable problem. Tissue cultures can become contaminated at any stage of tissue culturing process .By average, there are 60 contaminated specimens daily identified by manual visual inspection by the laboratory technician.

Fig.1. Map of Southern Leyte

This study is created for a contamination grading system for abaca tissue culture specimen using Naïve Bayesian classification. By utilizing Naïve Bayesian classification algorithm in tissue culture laboratory it will increase the accuracy rate of early prediction of contaminated specimen preventing the infected specimen from infecting other specimen in the laboratory.

Moreover, this study will augment the abaca rehabilitation

Manuscript submitted on December 30, 2016.

R. D. Malangsa is a doctorate student taking up Doctor of Information Technology at Cebu Institute of Technology-University, Cebu City Philippines

E. A. Maravillas is a full professor and Chairperson, of Graduate Program at Cebu Institute of Technology-University.

effort of Department of Agriculture thru the combined initiatives of LGU and SLSU. This study aimed also to determine the accuracy, precision, and specificity of the Naïve Bayesian classifier on the in vitro specimens.

There have been many studies before utilizing Naïve Bayesian algorithm on disease prediction especially for crops but the study on in vitro specimens has not given enough focused yet.

II. RELATED STUDIES

In fruit grading study, features are extracted from the fruit to classify fruit disease.[4] Extracted features like color, texture and shape of the fruit integrated with the Naïve Bayesian classifier built on histogram matching and region based approaches.

Bayesian can be applied in expert system for diabetes prediction while there is a possibility of existing attribute values and new samples and can be measures in terms of probability class labels [5]. Instead of classifying the traditional testing data with training data, the initial training data to the optimal process is forwarded, to extract the optimal data set; on that optimal dataset Naive Bayesian classifier is applied.

Naive Bayes Classifier was able to classify the agricultural land soil data resulted to 100% classification of instances [6]. Others [7], [8] have demonstrated the value of image processing in inspecting and grading the quality of agricultural and food products.

While an automated system for the disease detection and grading in pomegranate plant was proposed [8]. The techniques used includes color segmentation based on linear discriminant analysis, contour curvature analysis and a thinning process, which involves iterating until the stem becomes a skeleton.

Feature extraction methods and classification techniques are applied systematically in the attempt to solve the problem in plant disease classification [9]. The classification algorithm has feasibly identified the two diseases in the banana. Features that have been selected that work best for this application are when H and S color components are combined with the five shape features that were chosen as most important

III. METHODOLOGY AND PROCEDURE

Prototyping methodology is a software development process which allows developers to create portions of the solution to demonstrate functionality and make needed refinements before developing the final product.

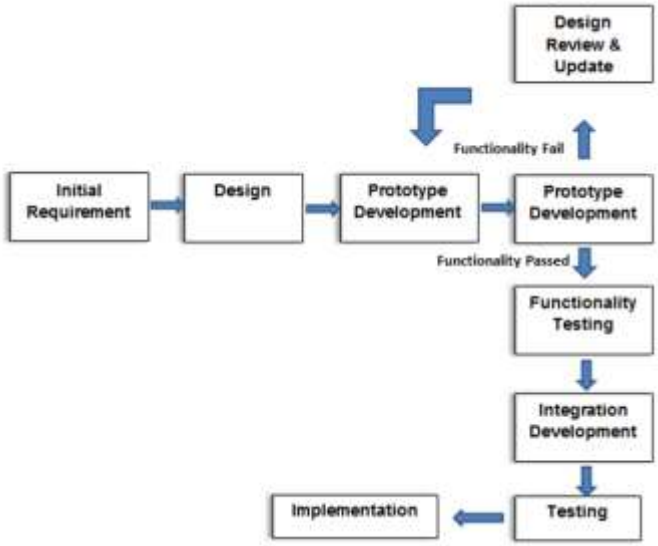

Fig. 2. Prototyping methodology

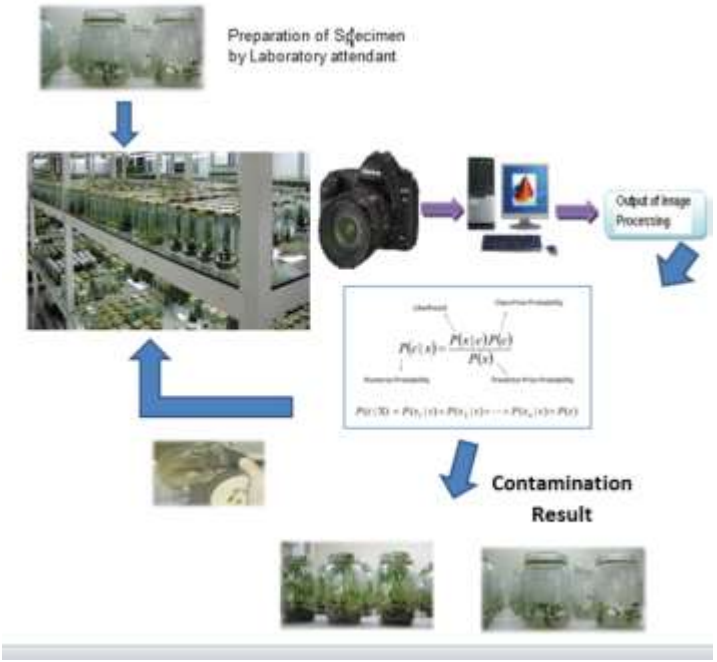

Fig. 3. The architectural lay-out of the developed system

The developed system was made possible with the utilization of raspberry pi microcontroller, OpenCV, Zigbee wireless data transmission, and 12 megapixel camera

1. Capture Specimen

A camera of 12 megapixels was used to capture both healthy and diseased images. In order to capture clear images with descriptive details, the camera was kept in both horizontal and vertical resolution of 72dpi (dots per inch). The distance of the camera from the specimen is 32 cm. The elevation of the camera is 23 cm. The angle of the camera is 95^0.

2. Feature extraction

RGB color space (Red, Green, and Blue) is the combination of the primary colors of red, green, and blue, which is used by a computer monitor or television [14].The color results from a combination of three colors and each - each have a value of 8 bits of red, 8 bits of green, and 8 bits of blue. Mixture of the three primary colors balanced with porpoise will produce shades of gray. If three fully saturated colors, it will produce white. Then an RGB mean values was displayed and recorded in the database.

The captured images from video streams may contain many objects especially in the background and working with such images leads to incorrect results. These images were cropped however, cropped images then had a white background with pixel values of 255 and working with the whole image also brings inappropriate results too. To avoid this challenge a mask was applied onto the image in order to obtain the useful segment.

Region of interest is used to define the area to the specimen.This converts an image from video streams into binary, thus indicating the segmentation of specimens from the background. Separate maskings for whitish features also were converted into binary image that mostly found in the contaminated specimens.

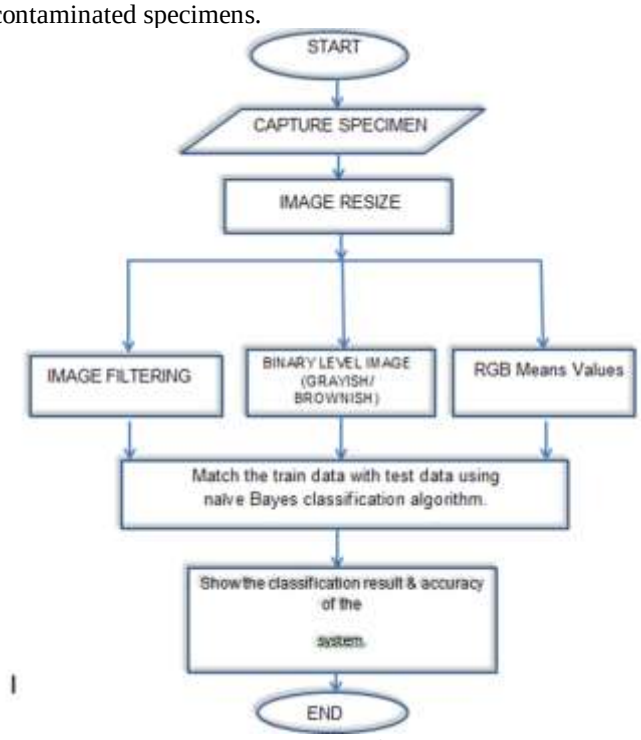

Fig. 4. Flowchart of the system

3. Naïve Bayesian Classification

Naïve Bayesian Classification is commonly known as a statistical means classifier. Based on Bayes' Theorem, and uses probabilistic analysis for effective classification. [10] It give more accurate results in less computation time when applied to the large data sets consisting of hundreds of images. The formula for Naïve Bayes classifier is: P (H | E) = P (E | H) x P(H) / P(E). The basic idea of Bayes rule is that outcome of a hypothesis or an event (H) can be predicted based on some evidences (E) that can be observed.

The advantages of Naive Bayes are [13]:

• It uses a very intuitive technique. Bayes classifiers, unlike neural networks, do not have several free parameters that must be set. This greatly simplifies the design process.

• Since the classifier returns probabilities, it is simpler to apply these results to a wide variety of tasks than if an arbitrary scale was used.

• It does not require large amounts of data before learning can begin.

• Naive Bayes classifiers are computationally fast when making decisions.

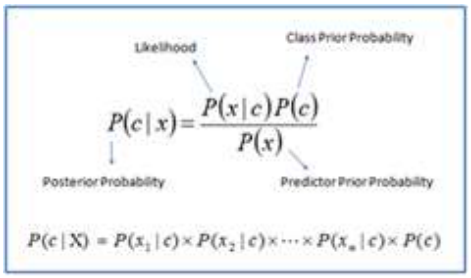

Fig. 5. The naïve Bayesian formula

Where;

- P(c|x) is the posterior probability of class (target) given the predictor (attributes).
- P(c) is the prior probability of class
- P(x|c) is the likelihood which is the probability given class
- P(x) is the prior probability of predictor.

$P(c|x) = P(x_1|c) \times P(X_2|c) \times \ldots \times P(x_n|c) \times P(c)$

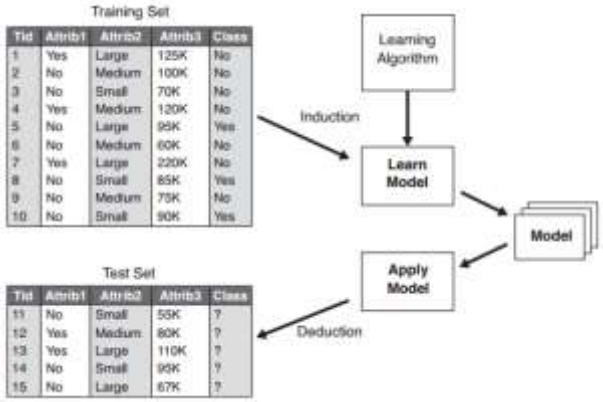

Fig. 6. General approach for building as classification model

Table 1 and Table 2 below represent the values of RGB mean and Binary (Whitish) respectively in categorical form. These values were approved by laboratory technician after a series of trial and errors experimentation (calibration).

TABLE 1. RGB MEANS VALUES

Level	Values
Level One	<= 0.3065039062
Level Two	> 0.3065039062 to <= 0.6130078124
Level Three	> 0.6130078124 to <= 0.9195117186
Level Four	> 0.9195117186 to <= 1.2260156248
Level Five	>1.2260156248 to <= 1.532519531
Level Six	>1.532519531 to <= 1.8390234372
Level Seven	> 1.8390234372

TABLE 2. BINARY VALUES FOR WHITISH

Level	Values
Level One	<= 50
Level Two	> 50 to <= 100
Level Three	> 100 to <= 200
Level Four	> 200 to <= 500
Level Five	> 500 to <= 1000
Level Six	> 1000 to <= 2000
Level Seven	> 2000

4. Training and Testing

The study was conducted at the Abaca Tissue Culture, at present the laboratory technician manually inspect the in vitro specimens for contamination, by average there are 500 specimens prepared daily. This technique is an integral component of the current replanting program. In addition, tissue culture has become standard for commercial plantations in recent years, primarily because of the advantage of starting with disease-free planting material [11].

Contamination in the specimen is caused by fungi and bacteria, the visual appearance of these contamination is the whitish cottony appearance in the medium caused by fungi, while a black dots caused by bacteria.

Frequently encountered bacterial and fungal contamination especially in laboratories of micro propagation posed a considerable problem. Tissue cultures can become contaminated at any stage of tissue culturing process. Fungus may arrive with an explant, or airborne, or enter a culture. The presence of people and contamination levels may get high when the area is heavily populated.

The principal microbial contaminants frequently reported in plant in vitro cultures are bacteria and fungi [12]. The main fungal contaminants frequently observed in plant tissue cultures are *Alterneria tenius*, *Aspergillus niger*, *Aspergillu fumigatus* and *Fusarium culmorum*

TABLE 3. CHARACTERIZATION AND IDENTIFICATION OF FUNGAL CONTAMINANTS OF TISSUE-CULTURED ABACA (MUSA TEXTILES NEE).

Contaminants	Cultural characteristics	Morphological characteristics
Aspergillus sp.	Colonies are flat, circle, Filamentous, velvety, wolly or cottony texture. Colony color is gray to green at center with a white border. The reverse is yellow to Pale yellow.	Conidiophores bear heads, long and hyaline that terminates in bulbous heads while conidia are globose to subglobose and usually yellowish green and dark brown
Chrysosporium sp.	Colonies are semi elevated, circle, fairly rapid grower, smooth. Colony color is white to off-white. The reverse is white to off-white color	Produced septate, hyaline hyphae. Conidia often appeared to be minimally differentiated from the hyphae and may appear to form directly on the hyphae. Conidia more often formed at the ends of simple or branched conidiophores of varying lengths. Conidiophores were ramified, forming tree-like structures

During the training phase, there were 200 specimen of which divided into 6 sets. `140 for healthy specimens, 15 for slightly contaminated, 15 for moderately contaminated, 15 for heavily contaminated, and 15 for critically contaminated specimens

Due to the huge amount if specimens and the varied characteristics of contaminations in the abaca tissue culture, the researchers had come up with the identified classes contaminants as agreed with the certified laboratory technicians.

After the training, the five classes for contamination/output as identified by the laboratory technician as follows: (a) contamination free, (b). slightly contaminated, (c) moderately contaminated, (d) heavily contaminated, and (e) critically contaminated.

For testing phase, there were new 98 specimens tested, in this phase the decision of the classifier and the decision of the laboratory technician were evaluated in actual contamination grading.

Fig. 7. Preparation of abaca tissue culture specimen (in vitro)

Fig. 8. The specimen inside the tissue culture laboratory

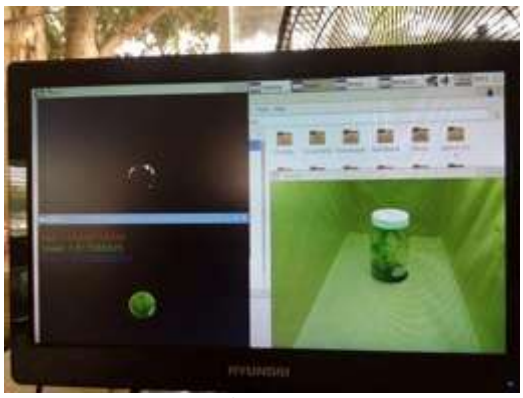

Fig. 9. The monitor showing the feature extraction during the training phase

Fig. 10. The system inside the abaca tissue culture laboratory

IV. RESULTS AND DISCUSSION

A confusion matrix [15] contains information about actual and predicted classifications done by a classification system. It shows the number of correct and incorrect predictions made by the classification model compared to the actual outcomes (target value) in the data.

Performance of such systems is commonly evaluated using the data in the matrix.

TP (True Positive), when the outcome is correctly classified as "yes" or "positive", when it is "yes" or "positive". **TN** (True Negative) is when the outcome is correctly classified as "no" or "negative" , when it is no "no" or "negative". On the other hand, **FN** (False Negative) is when the outcome is incorrectly classified as negative when it is in fact positive. **FP** (False Positive) when the outcome is incorrectly classified as "yes" , when it is in fact a "no" or "negative".

TABLE 4. THE CONFUSION MATRIX

		Testing (Target)				
		CF (Contamination Free)	SC (Slightly Contaminated)	MC (Moderately Contaminated)	HC (Highly Contaminated)	CC (Critically Contaminated)
Training (Model)	CF (ContaminationFree)	68	2	0	0	3
	SC (Slightly Contaminated)	0	2	0	0	0
	MC (Moderately Contaminated)	0	0	3	0	0
	HC (Highly Contaminated)	1	0	1	5	3
	CC (Critically Contaminated)	0	0	0	0	10

It can be seen in the confusion matrix the TP values (in diagonal), in which the values are not so scattered.

Accuracy is calculated as the sum of correct classification divided by the total number of classification (the diagonal). During the testing with new 98 in vitro specimens, the overall accuracy of the system based on confusion matrix is **90%,** while the number of false positive is **10.** This is quite a high accuracy.

$$AC = \frac{a+d}{a+b+c+d}$$

Precision is the proportion of the predicted positive cases that were correct, as calculated using the equation;

$$\text{Precision} = TP_A / (TP_A + FP_A)$$

Recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation;

$$\text{Recall} = TP_A/(TP_A + FN_A)$$

Specificity is the proportion of actual negatives classes which are correctly identified;

$$\text{Specificity} = TN_A/(FP_A + TN_A)$$

TABLE 5. THE RESULT OF THE PERFORMANCE OF THE CLASSIFIER

Class	True Positive	True Negative	False Positive	False Negative	Precision	Recall/Sensitivity	Specificity
CF (Contamination Free)	68	20	1	5	98.5%	93.15%	95.23%
SC (Slightly Contaminated	2	86	2	0	50%	100%	97.72%
MC (Moderately Contaminated)	3	85	1	0	75%	100%	98.83%
HC (Highly Contaminated)	5	83	0	5	100%	50%	100%
CC (Critically Contaminated	10	78	6	0	62.5%	100%	92.85%

In the CF(contamination free) class, the precision is 98.5% and recall is 93.15%. This means that for precision, out of the

times CF class was predicted, 93.15% of the time the system was in fact correct. And for recall, it means that out of all the times CF class should have been predicted by 93.15% of the labels were correctly predicted.

It is also noticeable that in highly contaminated class (HC), the system was able to come up with precision percentage of 100% and specificity percentage of 100%.

V. CONCLUSION

With a high overall accuracy of 90% of the system in contamination grading, this research has proved that there is a constitent and more accurate way of detecting contamination in the tissue culture specimens rather than the naked eye inspections. Moreover, the percentage generated in precision,recall, and specificity also indicate the very good performance of the classifier. The study indicated that Naive Bayes has good potential for identifying contamination grading accurately that mainly causes by fungi in abaca tissue culture laboratory. Though some factors needs to be considered during the training phase like the distance of the camera to the specimens, the amount of light, the angle of the camera, and memory of the microcontroller. The system cannot match the precision and accuracy of the human eye, but the speed and the cost at which they work can be easily be overcome.

REFERENCES

[1] Di Palma, D., Bencini, L., Collodi, G., Manes, G., Chiti, F., Fantacci, R., "Distributed Monitoring Systems for Agriculture based on Wireless Sensor Network Technology", Department of Electronics and Telecommunications, University of Florence Via di Santa Marta 3, 50019 Florence Italy

[2] Galo, E., Pedroso, J., "Module 5: Tissue Culture of Banana and Abaca", Western Mindanao State University, Zamboanga City 2009

[3] Erna S. Gorne (30 October 2006). "Bunchy top virus in Southern Leyte scales down abaca production". PIA Information Services - Philippine Information Agency. Archived from the original on 5 May 2007.

[4] U. Solanki, U.K. Jaliya, and D.G. Thakore (2015), " A survey on detection of disease and fruit grading', *International Journal of Innovative and Emerging Research in Engineering*, vol. 2, no. 2, pp. 109-114
Available: http://www.ijiere.com

[5] A. Ambica, S. Gandi, and A. Kothalanka (October 2013), "An efficient expert system for diabetes by naïve Bayesian classifier", *International Journal of Engineering Trends and Technology*, vol. 4, no. 10. pp. 4634-4639
Available: http://www.ijettjournal.org

[6] P. Bhargani and S. Jyotchi (August 2009),"Applying naïve Bayes data mining technique for classification of agricultural land soils". *International Journal of Computer Science and Network Security*, vol. 9, no. 8, pp. 117-122
Available: http://www.ijcsns.org

[7] N.V.G. and H.K.S. (May 2010), "Quality inspection and grading of agricultural and food products by computer vision", *International Journal of Computer Applications*, vol.2, no. 1

[8] S, Sannakki, V. Rajpurohit, V. Nargubnd, and R. Arun Kumar.(2011), "Hybrid Intelligent system for automated pomegranate disease detection and grading" , *International Journal of Machine Intelligence* , vol. 3, no. 2, , pp-36-44

[9] G. Owomugisha, J.A. Quinn, and E. Mwebaze (2014), "Automated vision-based diagnosis of banana bacterial wilt disease and black sigatoka disease', *International Proceedings of the 1st International Conference on the use of Mobile in ICT in Africa* 2014. Pp. 1-5, December 9-10, 2014, Stellenbosch, South Africa

[10] S.Dhanodharam (May 2014), " Liver disease prediction using Bayesian classification", Proceeding on 4th National Conference on Advance Computing, Applications and Technologies, special issue, pp. 1-3

[11] Raymundo, A., Bajet, N.," Epidemiology and Integrated Management of Abaca Bunchy Top in the Philippines". Department of Plant Pathology, University of the Philippines, Los Baños, Philippines.

[12] J.S. Cobrado and A.M. Fernandez (2016), "Common fungi contamination affecting tissue-cultured abaca (musa textiles nee) during initial stage of micro propagation", Asian Research Journal of Agriculture 1(2): 1-7
Available: http://www.journalrepository.org

[13] M. K. Stern, J. E. Beck, and B. P. Woolf, "Naïve Bayes Classifiers for User Modeling,"
Available:
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.118.979

[14] R. Rulaningtyas, A. B. Suksmono, L.R Mengko, and P. Saptawati (December 2012), "Color segmentation using Bayesian method of tuberculosis bacteria images in ziehl-neelsen sputum smear", *Conference Paper*
Available: https://www.researchgate.net/publication/259636216

[15] R. Kohavi and F. Provost (1998)," Special issue on applications of machine learning and the knowledge discovery process", Machine Learning, Kluwer Academic Publishers, Boston, Manufactured in The Netherlands 30, 271-274
Available: http:// robotics.stanford.edu/~ronnyk/glossary.html

Rhoderick D. Malangsa. Currently an assistant professor at Southern Leyte State University, Sogod Southern Leyte. He obtained his BS Computer in Science degree at STI College, Ormoc City in 2004, finished with the degree of MS in Information Technology at Southern Leyte State University, Sogod Southern Leyte in 2012. He also earned 32 units in MS in Teaching Computer Science. He is currently on his dissertation writing pursuing the degree Doctor of Information Technology at Cebu Institute of Technology-University, Cebu

Elmer A. Maravillas. Is a Full Professor. Currently He is Chair, Graduate Program and BSComputer Science at Cebu Institute of Technology-University. He finished with the degree of BS Mechanical Engineering in 1981 at Western Mindanao State University, Zamboanga City, MS Computer Science in 1998 and PhD in Mechanical Engineering in 2006 from De La Salle University-Manila. Research interests include modeling and simulation, fuzzy logic systems, neural networks, intelligent systems and expert systems.