CAREdio: Health Screening and Heart Disease Prediction System for Rural Communities in the Philippines

Lean Karlo S. Tolentino[1] , John Erick L. Isoy[2], Kayne Adriane A. Bulawan[3], Mary Claire T. Co[4]
Caryl Faye C. Monreal[5], Ian Joshua W. Vitto[6], Maria Victoria C. Padilla[7], Jay Fel C. Quijano[8]
Romeo Jr. L. Jorda[9], Jessica S. Velasco[10]

Department of Electronics Engineering, Technological University of the Philippines, Manila, Philippines[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
University Extension Services, Technological University of the Philippines, Manila, Philippines[1]

Abstract—**Cardiovascular diseases cover a large quantity of worldwide disease load, setting it to top leading cause of death. In the Philippines, given the rapid economic advancement and urbanization, the most vulnerable sector has not been impacted by this development. Data from the Philippine Statistical Authority (PSA) in 2016 revealed that of the country's total recorded deaths, six out of ten were medically unattended and of which the largest portion are from the rural population. Consequently, medical analysis is needed to perform effectively and precisely however, most developing countries have limited resources and lack medical expert for specialized field such as cardiologists. The proponents essentially seeks to address the issues Philippine health sector specifically in rural and remote populace by executing efficient and low-cost health screening and diseases prediction system using commercially available medical devices and machine learning algorithms for the prediction of three of the most heart diseases (Hypertension, Heart Attack, Diabetes). The system is composed of CAREdio mobile app, prototype hardware consists of different health sensors and devices, and a machine learning model that is applied to determine the user's individual probability of having a specific heart disease. The machine learning models used were trained using the data gathered from Rosario Reyes Health Center and Ospital ng Sampaloc (Sampaloc Hospital), both located in Manila City, Philippines. CAREdio achieves accuracy values over 0.80 for all diseases. The system can diagnose multiple cardiovascular diseases in a single app that will benefit people rural communities.**

Keywords—Cardiovascular diseases; health screening; disease prediction; mobile application; machine learning; rural population

I. INTRODUCTION

The Philippines is an archipelago of over 7,000 islands, with an entire land scale of approximately 300,000 square kilometers, and above 100 million inhabitants, placing it to the thirteenth most populated country in the world [1], these brings a genuine task in providing efficient and effective health care particularly for rural and remote population.

In spite of the rapid economic advancement and urbanization, the most vulnerable sector has not been impacted by this development. Philippine Statistical Authority revealed that of the country's total recorded deaths in 2016, six out of ten were medically unattended. The largest portion of medically unattended deaths are from the rural population. Only the National Capital Region (NCR) had a greater total of medically assisted deaths than unattended in general. It infers that compared to other regions; NCR has better access to quality health care [2].

Because of the maldistribution of facilities, health staff and specialists especially in rural regions, it has put the residents into a terrible situation. The limited number of health facilities relating to the growing population, and shortage of physicians add up to a low quality of health care. Likewise, based on a review by the Social Weather Station, a large portion of Filipinos, particularly the low-income households, choose to get treatment in a government-owned medical facility if a member of the family needs confinement and it is mostly because of affordability [3].

The world's leading cause of deaths are non-communicable diseases (NCDs) with 41 million individuals die every year, which equates to 71 percent of worldwide deaths. [4]. In 2016, approximately 17.9 million people died from heart illnesses, indicating the 31 percent total number of deaths. Eighty five percent of these deaths are because of heart attack and stroke. In the Philippines, NCDs are increasing quickly, seven of the ten top causes of death are non-communicable in nature. Most of the NCDs mortality cases are diabetes, chronic obstructive pulmonary disease, cardio-vascular diseases, and cancer [5].

The fast evolution of computational capacities has given present-day techniques disease prediction. Techniques that can reveal clinically important and relevant information concealed in the vast amount of data, which in this way, can support medical decision making. An artificial intelligence (AI) system can reduce medical and diagnostic mistakes, which in clinical practice is inevitable. Additionally, an AI system may collect valuable information from a huge population of patients to assist to make uninterrupted analysis for notifying health risks and calculating health findings. Machine learning algorithms are valuable for discovery complex patterns in large data.

The researchers principally pursue to address the issues in Philippine health sector specifically in rural and remote populace by executing efficient and low-cost health screening and diseases prediction system using commercially offered medical devices and machine learning algorithms. The specific

objectives of the study are: (1) to use commercially offered medical sensors and screening devices that can collect and read health parameters including heart rate (HR), body mass index (BMI), blood glucose (BG), oxygen saturation (SpO2), blood pressure (BP), uric acid level, and cholesterol; (2) to use the machine learning models such as Support Vector Machine, Naive Bayes Classifier, Random Forest, Logistic Regression, and Neural Network; and to predict the individual probabilities of a person having specific heart diseases; and (3) to evaluate the overall system in terms of accuracy and efficiency through comparison tests between the prediction engine and standard tests.

Data gathering for machine learning algorithm implementation was conducted at Rosario Reyes Health Center in San Andres, Manila, and Ospital ng Sampaloc, a District Hospital also in Manila. The collected data are from patients who have heart diseases. From this, the researchers were able to come up with three most common non-communicable diseases namely: heart attack, hypertension, and diabetes. Some other data collected were compacted into a category called as others.

The prototype was deployed at an experimental scale in Barangay Medicion II-B in Imus City, Cavite, Philippines. The prototype is composed of weight sensor, height sensor, heart rate sensor, glucometer, cholesterol meter, uric acid meter, oxygen saturation sensor, and sphygmomanometer for blood pressure, microcontroller, thermal printer, and android tablet installed with CAREdio application. The testing of patient is done in less than 10 minutes, depending on the patient. At first, the patient would have to enter his name, age, and sex in the application before proceeding to the actual testing. The patients will also have to answer several medical questionnaires integrated in the application that would help in prediction of their individual probabilities of having diseases.

II. CONCEPTUAL LITERATURE

A. Non-Communicable Dieseases

Non-Communicable Diseases (NCDs) are the premier causes of deaths in the country. NCDs such as pulmonary diseases, cardiovascular diseases, diabetes, and cancer.

Cardiovascular Disease (CVD) may be a complex class for certain diseases that can influence the heart and blood vessels. The CVD affects the structure or functions of the heart which include inherent cardiopathy, artery illness (narrowing of the arteries), irregular heart rhythms (arrhythmias), heart valve disease, cardiac failure, and muscle tissue illness. These diseases occur because of imbalance in blood pressure and pulse rate. Clear addition to the CVD disease is the increase in blood pressure, smoking, sterol (diabetes polygenic disorder polygenic disease) [6].

Some variables known to have a substantial percentage of risk for CVDs include age, gender, heart rate, level of cholesterol, level of blood pressure and BMI.

Diabetes is deemed as one of the most pervasive illness. It characterizes a group of high - glucose metabolic illnesses in the blood basically because of inappropriate ingestion of insulin. Preserving the balance between hypoglycemia (low glucose level) and hyperglycemia (high glucose level) is the job of insulin.

There are three types of diabetes: type 1 it the condition where the pancreas is not producing insulin, and type 2 is where the pancreas creates insulin but is not adequate to sustain the stability.

Heredity, obesity, hypertension, aging, unbalanced diet, lack of physical activity, and because of the nature of community are some of the factors that contribute to having type 2 diabetes. In type 2 diabetes, cardiovascular disease is critical. An additional type where the amount of glucose is increasing during pregnancy is called is called Gestational Diabetes [7].

B. Health Parameters

To get a reliable picture of one's wellbeing, predict pattern and screening of health, several health factors are being accessed. Table I shows the health parameters, values, and remarks which is to be used in arranging the data collected and ran using machine learning algorithms [5], [8 - 12].

C. Machine Learning Algorithms

The study of computer algorithms that develops naturally through repetitive experience is called machine learning algorithms (MLA). It is viewed as a subgroup of artificial intelligence. Machine learning algorithms constructs a mathematical model established on sample data, known as "training data", so as to make predictions or decisions without being openly coded to do so [13]. In this study, we utilized 5 of most common machine learning models:

1) An array of layers consists of many very interconnected parts (i.e. neurons) doing an activation is how neural network is designed. One or more concealed and the output layers are gradually generated when there is an interaction in the input layers [14].

2) Random forest is a regression and classification model that builds decision trees for every attribute; modifies the overfitting to its training set; keeps outliers of missing quantities by doing steps in analysis, pre-processing data. It's a method of creating a specific model anywhere poor models are associated. The random forest is made up of the numerous decision trees connected to the forest [15].

3) Support Vector Machine can be used for classification and regression. Though, this model is usually used in the classification applications. The SVM definition focuses on finding a dimension (straight line for 2D; plane for 3D) which offers the best path to split the groups [16].

4) Naive Bayes Classifier is a classification method that is established on Bayes' Theorem with a notion of autonomy among predictors. Naive Bayes presumes in simple terms, that the existence of an element in a class is unrelated to the presence of any other element [17].

5) Logistic regression is a model that uses logistic function to build a binary dependent variable with two possible values. It can be compatible to a dependent variable. Binary variables, 1 and 0 indicating they succeed or fail [18].

TABLE I.   LIST OF HEALTH PARAMETERS

No.	Attribute	Health Parameter	Value	Remarks
1	age	Age	-	-
2	sex	Sex	0 1	Female Male
3	bmi	Body Mass Index	0 (18.5 – 22.9) 1 (23 – 24.9) 2 (25 – 29.9) 3 (=> 30)	Normal Overweight Pre-obese Obese
4	maxhr	Max Heart Rate	0 (60-90 bmp) 1 (>90 bpm)	Normal Abnormal
5	bos	Blood Oxygen Saturation	0 (95%-100%) 1 (<94%)	Normal Hypoxemia
6	fbs	Fasting Blood Sugar	0 (<100 mg/dl) 1 (101-125 mg/dl) 2 (>126 mg/dl)	Non-diabetic Pre-diabetes Diabetic
7	bpsys	Blood Pressure (systolic)	0 (<120) 1 (120-139) 2 (>140)	Normal At Risk High
8	bpdio	Blood Pressure (diastolic)	0 (<80) 1 (80-89) 2 (>90)	Normal At Risk High
9	chllvl	Cholesterol Level	0 (<200 mg/dl) 1 (201-239 mg/dl) 2 (=< 240 md/dl)	Desirable Borderline High
10	sm	Smoking	0 1	False True
11	famhis	Family History	0 1	False True
12	alchl	Alcohol	0 1	False True
13	stress	Stress	0 1	False True
14	hrtbrn	Heart Burn	0 1	False True
15	dt	Diet	0 1	False True
16	shrtnssbrth	Shortness of Breath	0 1	False True
17	exrcs	Active Exercise	0 1	False True
18	cp	Chest Pain	0 1	False True
19	num	Disease	0 1	False True

III. RELATED WORKS

Disease prediction using artificial intelligence is recently getting international acknowledgment. Presented the large scope of data mining in enhancing health care, this part discusses the methods used in determining and forecasting NCDs.

The predictive model calculates the future results based on past records recovered from database. This kind of model is used by many organizations that try that uses AI to try and understand the connections between the information characteristics. Researchers have suggested and used a variety of techniques and machine learning algorithms which they have implemented in some medical applications. Some of which are as follows:

In [19], the researchers created an Android based app that provides choices as to what patients should do when they experience a specific symptoms, may it be doing a medication at home, or going to a clinic to be checked by a doctor right away. Chaining inferences forward is the method used in this study. Clinical results of a disease can be identified with the use of application and Android phone as the medium through the consultation procedure or answering the questionnaires integrated in the system. After doctors had accessed the result of the general diagnosis symptoms, the concluded that treatment it provides are effective.

Similar study [20] projected a new ear - worn system for intermittent long - term ECG QRS duration analyzing to control the difficulties of recent wearable ECG technologies for example inconvenience of use. To enhanced wear experience, the ECG electrodes was placed at the back of the ear. The

researchers applied a heartbeat validation SVM model, a heartbeat purification method, and a regression model to extract typical chest-ECG QRS from the ear's QRS durations – ECG.

In [21], the proponents discussed all what is called a Hierarchal Health Decision Support System which fuses clinical decision support system and wearable medical sensors (WMSs). Closed - loop and multi - tier hierarchical system supported by stable machine learning classes are incorporated in the system. The proponents proposed an automated application for disease diagnosis, which could investigate various related diseases. Accuracy in the classification of major diseases was achieved as a result of physiological data gathered from WMSs: hypothyroid (95 percent), type 2 diabetes (78 percent), arrhythmia (86 percent), renal pelvis nephritis (94 percent), and urinary bladder disorder (99 percent).

In [7], blood glucose levels using the capacitance method were successfully measured by the proponents. The Artificial Rain Algorithm (ARA) was employed as a framework for the internal discovery of the finest value from the information collected via repeated examination of the capacitance value. The prototype effectively demonstrated the accuracy of 96.92% as contrasted to the laboratory test results. The researchers recommended the use of additional clustering methods to further develop the potential of ARA as an enhancing and analyzing methods.

In [22], the researchers used MLA to predict heart disease, namely: Decision Tree, SVM, and Naive Bayes. The principal component analysis was used to reduce the attributes number. After the size of the dataset has been reduced, SVM beats other two models. To analyze the accuracy of the algorithms the researchers used WEKA data mining method. The algorithms compare the accuracy of the classifier based on its time carried to construct model, ROC Area and mean absolute error. It concludes that the use of Max ROC Areas is in prediction performance was brilliant.

The study in [23] offered the evaluation and application of the given MLAs which was applied in R programming in producing CVDs prediction. Also, web-based application was also created which provides module for the attributes, test results, references, and graphs for entry. The top accuracy was obtained from logistic regression.

Artificial neural network was used in [24] for detecting circulatory diseases through analyzing fingernails. The developed model was tested on 6 patients' fingernail images. Their diagnosed diseases were all matched with the proposed prediction system.

Meanwhile, potentiometric method was implemented in [25] for detecting hormonal imbalance through the females' saliva. The results from the proposed method were compared with the conventional blood test and were all matched.

In [6], a system that produces a prediction if a patient has cardiovascular disease, in terms of YES or NO. Unless the patient is susceptible to cardiovascular disease then the outcome would be YES and vice versa. In case of good results, the patient needs to consult a doctor for more opinion. The cardiovascular disease prediction program uses the MLA which gives the patient a predictive result which gives the patient status which leads to CAD.

IV. METHODOLOGY

This section will describe methodology followed in developing the CAREdio.

A. Block Diagram

The CAREdio system is comprised of two principal elements. The first item is the set of medical sensor and system safety parameters including and the standardized clinical questionnaires. The second half is for assessment and prediction engine.

The block diagram in Fig. 1 shows how the CAREdio works step by step. This comprises the parameter detection from the prototype kiosk which includes medical sensors and equipment, microcontrollers for the acquisition of data and the analysis, an android application where the user can enter their information and responses to the questions. Analysis of data and prediction of diseases will be done by using the MLA.

B. Machine Learning Implementation

Specific datasets of the diseases were collected to check and train the MLAs. The evaluation and selection of the accuracy and efficiency of the MLA in predicting heart disease shall be the deciding factor in evaluating which one is the best to use. The preferred MLA will then be used and be integrated to the CAREdio system for prediction of individual probability of having diseases.

C. System Flow Chart

Fig. 2 illustrates the flow chart of the CAREdio system. First, the system will prepare and wait for the patient's question to display from the application that we will be using. It will then start to gather the patient's data by filling up the information needed for new users. After logging in, medical sensors will operate to detect patient's health parameter such as height, weight, BMI, blood pressure rate and medical questions will be shown in the android application. The data gathered from the patient will be processed for detecting the probability of having a heart disease. When the results are done, it will automatically be saved in database and the patient has an option whether to print their results or not.
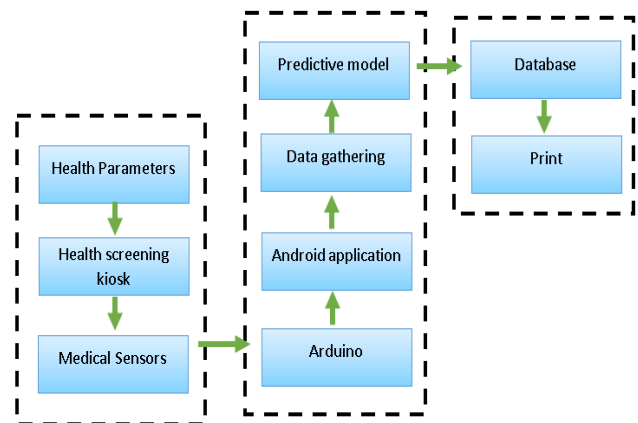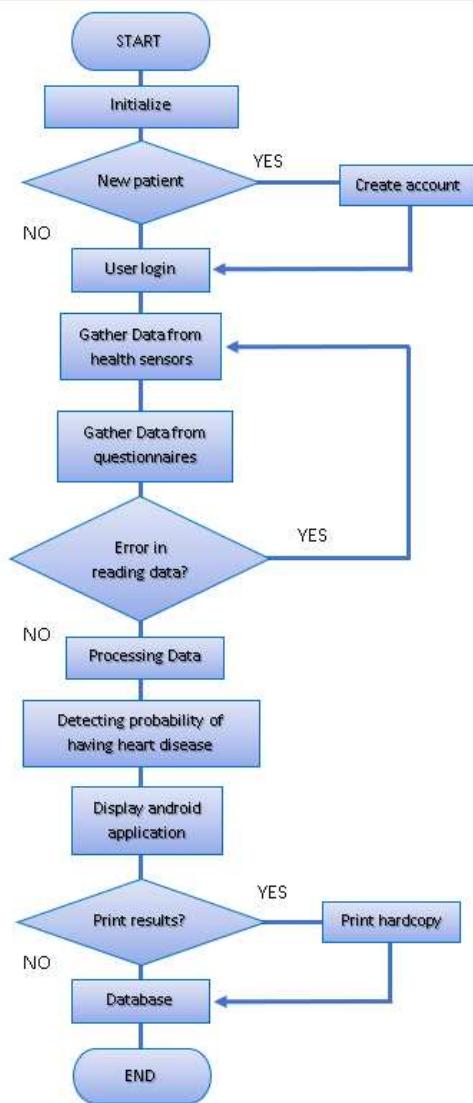

Fig. 1. Block Diagram of the System.

Fig. 2.   System Flow Chart.

Fig. 4 shows the database in android application. The data of the patient were forwarded to the database using SQLite and saved in csv file format. The database consists of timestamp, name of patients, age, health parameters and result percentage. SQLite is an open source relational database that can store and access data across multiple storage engines. It can also duplicate data and sectioning tables for better durability and performance.
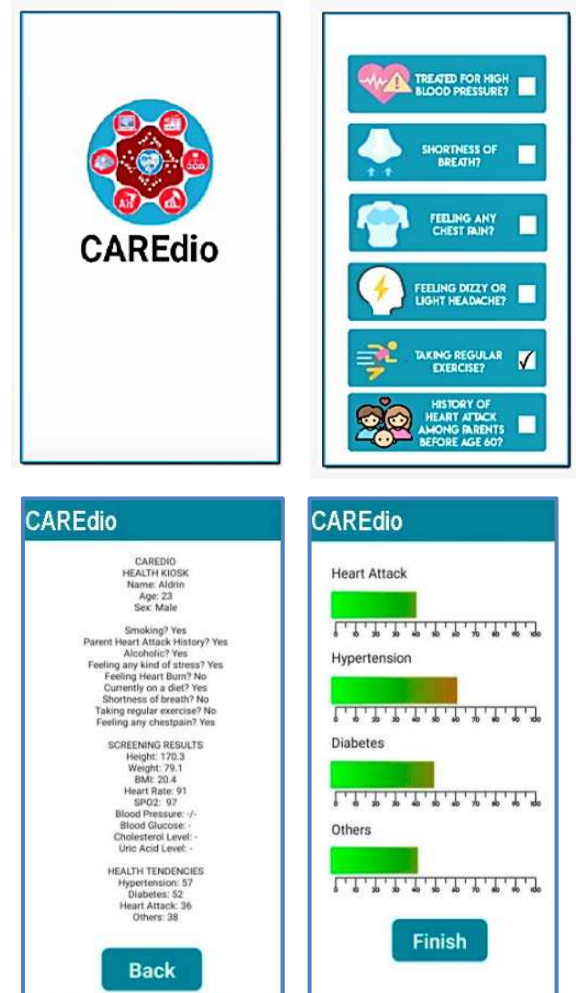

Fig. 3.   From Left, Screenshots of CAREdio Application, (a) Home Page, (b) Input Parameter, (c) Medical Questionnaire.

D. Mobile Application Design

This section presents all of the tabs included in the Android application. Every tab view has a different function for the system which include manual input of patient's information, obtaining of parameters from the prototype, questionnaires, patient and record information, and display of results.

The mobile application design is shown in Fig. 3. On the app interfaces, the user will enter personal data like name, age and gender; after that, they will have to use the prototype for health screening to gather parameters, answer the standard clinical questionnaires. The patients can view the data collected included the prediction percentage, they will also be given also get a hardcopy result. The section for admin is where the data base can be viewed and extracted in .CSV file.

It shows the display of patients result in android application. The results obtained from medical devices and sensors, responses in the questionnaire, and patient information are shown. The probability of having those diseases are also displayed by percentage.

Fig. 4.   Database in Android Application Sample.

E. Hardware Prototype

Fig. 5 shows the proposed hardware structure for the health screening kiosk. It is composed of microcontrollers, medical sensors and equipment, Android tablet, and printer. While the actual prototype is shown in Fig. 6. The height is 4 feet (which be expanded to 6 feet max for the height sensor), while the width spans 1.5 feet. Height sensor (HC-SR04) is positioned at the top of the prototype kiosk. For the middle left, shows the additional sensor, which is MAX30105 for heart rate and blood oxygen saturation. The drawer in the middle is the storage for devices like digital blood pressure; the 3-in-1 test kit for blood glucose, cholesterol level, and acid saturation; Arduino Mega and Raspberry Pi microcontrollers. The device has an aluminum door for a weighing scale located at the lower part of prototype. Thermal printer and 7 inches Android tablet were also installed.
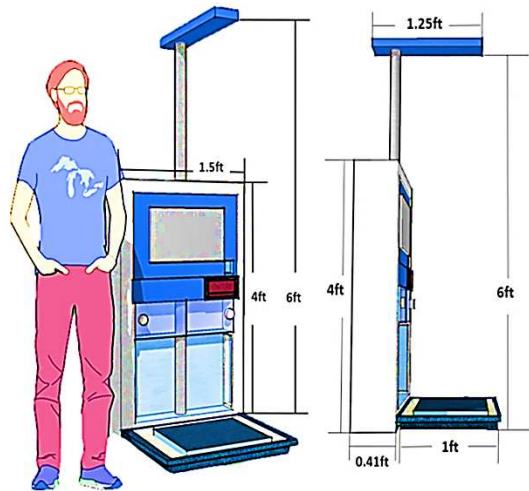

Fig. 5.   Proposed Prototype.

Fig. 6.   Actual Prototype.

Health parameter scores from health sensors and devices, and the answers from medical survey serves as input for the system. The patient will stand on the platform of the prototype for the detection its height and weight, and they are to place their index finger in the pulse oximeter sensor for the blood oxygen saturation and heart rate reading. The blood pressure, cholesterol level, uric acid level, and blood glucose level, will be collected from medical devices and will be manually recorded in the mobile application along with answering the medical questionnaire on it. When all inputs are entered in the system, the prediction of cardiovascular diseases will take place.

V.   ANALYSIS AND RESULTS

The CAREdio is health screening and disease prediction system that utilizes MLA to automate prediction of cardiovascular diseases, specifically (1) diabetes, (2) heart attack, (3) hypertension, and (4) others, a separate group labeled as for other cardiovascular diseases that is currently not included by our system. The system can provide the individual probability of having cardiovascular diseases by displaying each percentage output. CAREdio is projected to meet the standards of health associations, hospitals, and clinic tests with convenience and efficient cost.

The system is not fully automated, some medical devices are not integrated in microcontroller when getting health parameters, it also needs a manual encoding to generate in mobile application. It was created as a health screener for the stated health parameters and disease prediction, it does not give certain treatment, but it will give individual probability of having heart diseases based on the data gathered.

The Random Forest model is chosen for the calculation of the probability of having specific cardiovascular diseases, among the five machine algorithms including Logistic Regression Support Vector Machine, Naive-Bayes Classifier, and Neural Network, because it formed the highest percentage of accuracy when the data collected from hospitals was trained and tested.

A. Machine Learning Accuracy

Table II shows the summary of results of each algorithm. The highest percentage for machine learning is Random forest with average of 79.3% accuracy. Shows the summary of machine learning accuracy percentage for each disease. Random Forest got the highest accuracy rate for hypertension and heart attack.

Table III shows the accuracy score, precision, recall, f1-score, and mean absolute error. The model does well in most heart diseases with accuracy scores above 0.90, except for category "others", which has an accuracy score of 0.87. Precision recall are also above 0.90 for most groups, and f1-score is above 0.88 for all groups. The "others" category has low sensitivity which is to be expected because of the range of diseases that make up this group.

The following table numbers 4.7, 4.8, 4.9, and 4.10, below shows the project device's output comparing it to collected data from Rosario Reyes Health Center, and *Ospital ng*

Sampaloc [Sampaloc Hospital], both located in Manila City, Philippines.

The proponents tested the mobile application against our dataset by randomly choosing 20 patient's data as the "test set" and operating to this it in system to calculate the probability of diseases and the accuracy of the system.

Table IV shows that 20 out of 20 (100%) data of person with diabetes matched. The device coordinates with the results from the dataset collected which produced the individual probability of having diabetes.

TABLE II.   MACHINE LEARNING PERFORMACE PER DISEASE

	Logic Regression	Naive Bayes	Neural Network	Random Forest	Support Vector Machine (SVM)
Diabetes	83..33%	81.25%	77.08%	91.67%	87.50%
Heart Attack	80.25%	89.89%	82.72%	90.12%	80.25%
Hypertension	97.53%	90.12%	85.19%	98.77%	98.77%
Others	84.26%	82.41%	80.56%	87.96%	87.96%

TABLE III.   ACCURACY SCORE, PRECISION, RECALL, F1-SCORE, AND MEAN ABSOLUTE ERROR

	Accuracy score	Precision	Recall	F1-score	Mean Absolute Error
Diabetes	0.98	0.98	0.98	0.98	0.95
Heart Attack	0.91	0.91	0.90	0.89	0.78
Hypertension	0.98	0.99	0.99	0.99	0.99
Others	0.87	0.88	0.88	0.88	0.88

TABLE IV.   TRIAL FOR PATIENT'S WITH DIABETES

Patient No.	With Diabetes	CAREdio Result	Remarks
1	Yes	70% (Yes)	Matched
2	Yes	94% (Yes)	Matched
3	Yes	89% (Yes)	Matched
4	Yes	98% (Yes)	Matched
5	Yes	91% (Yes)	Matched
6	Yes	94% (Yes)	Matched
7	Yes	80% (Yes)	Matched
8	Yes	96% (Yes)	Matched
9	Yes	98% (Yes)	Matched
10	Yes	86% (Yes)	Matched
11	Yes	96% (Yes)	Matched
12	Yes	93% (Yes)	Matched
13	Yes	51% (Yes)	Matched
14	Yes	97% (Yes)	Matched
15	Yes	78% (Yes)	Matched
16	No	10% (No)	Matched
17	No	2% (No)	Matched
18	No	7% (No)	Matched
19	No	18% (No)	Matched
20	No	12% (No)	Matched

Table V shows that 20 out of 20 (100%) data of person with heart attack matched. The device coordinates with the results from the dataset collected which produced the individual probability of having heart attack.

Table VI shows that 20 out of 20 (100%) data person with hypertension matched. The device coordinates with the results from the dataset collected which produced the individual probability of having diabetes.

TABLE V.   TRIAL FOR PATIENT'S WITH HEART ATTACK

Patient No.	With Heart Attack	CAREdio Result	Remarks
1	Yes	88% (Yes)	Matched
2	Yes	60% (Yes)	Matched
3	Yes	72% (Yes)	Matched
4	Yes	84% (Yes)	Matched
5	Yes	83% (Yes)	Matched
6	Yes	80% (Yes)	Matched
7	Yes	82% (Yes)	Matched
8	Yes	83% (Yes)	Matched
9	Yes	91% (Yes)	Matched
10	Yes	91% (Yes)	Matched
11	Yes	82% (Yes)	Matched
12	Yes	86% (Yes)	Matched
13	Yes	87% (Yes)	Matched
14	Yes	92% (Yes)	Matched
15	Yes	84% (Yes)	Matched
16	No	08% (No)	Matched
17	No	19% (No)	Matched
18	No	15% (No)	Matched
19	No	07% (No)	Matched
20	No	22% (No)	Matched

TABLE VI.   TRIAL FOR PATIENT'S WITH HYPERTENSION

Patient No.	With Hypertension	CAREdio Result	Remarks
1	Yes	97% (Yes)	Matched
2	Yes	96% (Yes)	Matched
3	Yes	78% (Yes)	Matched
4	Yes	91% (Yes)	Matched
5	Yes	90% (Yes)	Matched
6	Yes	99% (Yes)	Matched
7	Yes	88% (Yes)	Matched
8	Yes	91% (Yes)	Matched
9	Yes	90% (Yes)	Matched
10	Yes	90% (Yes)	Matched
11	Yes	86% (Yes)	Matched
12	Yes	88% (Yes)	Matched
13	Yes	89% (Yes)	Matched
14	Yes	87% (Yes)	Matched
15	Yes	89% (Yes)	Matched
16	No	31% (No)	Matched
17	No	71% (Yes)	Mismatched
18	No	10% (No)	Matched
19	No	13% (No)	Matched
20	No	18% (No)	Matched

Table VII shows that 19 out of 20 (95%) data in "others" category matched. The device coordinates with the results from the dataset collected which produced the individual probability of having hypertension.

TABLE VII.   TRIAL FOR PATIENT'S WITH OTHER CVD

Patient No.	With Other CVD	CAREdio Result	Remarks
1	Yes	80% (Yes)	Matched
2	Yes	56% (Yes)	Matched
3	Yes	59% (Yes)	Matched
4	Yes	95% (Yes)	Matched
5	Yes	86% (Yes)	Matched
6	Yes	88% (Yes)	Matched
7	Yes	51% (Yes)	Matched
8	Yes	100% (Yes)	Matched
9	Yes	79% (Yes)	Matched
10	Yes	95% (Yes)	Matched
11	Yes	91% (Yes)	Matched
12	Yes	95% (Yes)	Matched
13	Yes	55% (Yes)	Matched
14	Yes	86% (Yes)	Matched
15	Yes	98% (Yes)	Matched
16	No	4% (No)	Matched
17	No	0% (No)	Matched
18	No	2% (No)	Matched
19	No	9% (No)	Matched
20	No	4% (No)	Matched

VI. CONCLUSION AND FUTURE WORK

The development of health screening and heart disease prediction was successfully executed for the three of the most prevalent heart diseases in the Philippines. In addition to this, the improvement and accuracy of the system was well implemented with Random Forest Classifier.

The system is screening the health parameters with the CAREdio prototype composed of different health sensors and devices and analyzing the data collected from the patients to provide the individual probability of having diseases.

CAREdio has good accuracy score, precision, recall, f1-score and AUC, and can easily be operated by a community health worker to screen health and predict diseases, and also help them to make decisions towards patient health risks. The system is low cost, as it only involves a CAREdio application and the hardware protype. Screening and predicting system such as the CAREdio can provide a significant role for community health workers and doctors to prevent disease and disability in remote and low resource areas. It produces results that make it closer to the real-life situations.

This study affirms that health centers or even government offices in rural areas need a program or application of software that can predict NCDs. It was necessary to have a huge amount of raw data to have proper classification rules which lead to appropriate prediction of NCDs. The forecast is helpful not only for physicians but also for neighborhood patients to test their health on constant basis. The system built must be constantly reviewed, and the programming should be optimized so that the technology can also be applied through several health centers and clinics.

For this study, the proponents would like to make the following recommendations to further improve the project. First is the use of additional medical sensors, this will increase the accuracy and sensitivity of the device. It is also recommended to use additional types of diseases for a wider scope which will make the project more useful. Thirdly, the researchers recommend that future researchers develop an application that will easily gather the readings of sensors and saving the file with .csv file format without the help of terminal emulation application. And lastly, the proponents would like to recommend increasing the number of samples to be validated since it affects the percentage accuracy of the device and the laboratory testing.

REFERENCES

[1] "Population of the Philippines (2019 and historical)," 2019. Available: https://www.worldometers.info/world-population/philippines-population/.

[2] P. L. Bersales, Deaths in the Philippines, 2016, Available: https://psa.gov.ph/content/deaths-philippines-2016.

[3] World Health Organization, "Non-communicable Diseases," 2018. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

[4] World Health Organization, "Cardiovascular Diseases (CVDs)", 2017. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[5] A. Mata and G. Jasul Jr, "Prevalence of Metabolic Syndrome and its Individual Features Across Different (Normal, Overweight, Pre-Obese and Obese) Body Mass Index (BMI) Categories in a Tertiary Hospital in the Philippines," Journal of the ASEAN Federation of Endocrine Societies, vol. 32, no. 2, pp. 117-117, 2017.

[6] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1275-1278, 2018.

[7] A. Cuello, K. Dizon, M. L. Fortu, J. G. Paguia, A. V. Tañola, J. Velasco, A. Aquino, and I. Valenzuela, "Blood Glucose Analysis Utilizing Artificial Raindrop Algorithm and Capacitance Method," in 2017 International Conference on Computer and Applications (ICCA), pp. 135-139, 2017.

[8] A. A. Jambora, "For Filipinos, high blood pressure guideline remains 140/90, not the US' 130/80," 2018. Available: https://lifestyle.inquirer.net/295738/filipinos-high-blood-pressure-guideline-remains-140-90-not-us-130-80/.

[9] Marcin, J., and Holland, K. (2017). Is My Blood Oxygen Level Normal? . [online] Available at: http://www.healthline.com/health/normal-blood-oxygen-level#oxygen-levels3.

[10] M. Ong, "Health numbers you need to know," 2016. Available: https://www.philstar.com/lifestyle/health-and-family/2016/11/22/1644371/health-numbers-you-need-know.

[11] World Health Organization, Technical Report #727, 1985; NCEP Report, 1998.

[12] C. Jimeno, "A Summary of the Philippines UNITE for Diabetes Clinical Practice Guidelines for the Diagnosis and Management of Diabetes (Part I: Screening and Diagnosis of DM)," 2011.

[13] T. Chakravorty, "How Machine Learning Works: An Overview," 2016.

[14] D. Team, "Introduction to Artificial Neural Network Model," 2018.

[15] D. Storey, "Random Forests, Decision Trees, and Ensemble Methods Explained," N.D.

[16] N. Bambrick, "Support Vector Machines: A Simple Explanation," 2016

[17] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," Journal of Intelligent Learning Systems and Applications, vol. 9, no. 01, 2017.

[18] G. F. Cooper, C. F. Aliferis, R. Ambrosino, J. Aronis, B. G. Buchanan, R. Caruana, M. J. Fine et al., "An evaluation of machine-learning methods for predicting pneumonia mortality,"Artificial intelligence in medicine, vol. 9, no. 2, pp. 107-138, 1997.

[19] I. M. Shofi, L. K. Wardhani, and G. Anisa, "Android application for diagnosing general symptoms of disease using forward chaining method," in 2016 4th International Conference on Cyber and IT Service Management, pp. 1-7, 2016.

[20] Q. Zhang, D. Zhou, and X. Zeng, "Hear the heart: Daily cardiac health monitoring using Ear-ECG and machine learning," in 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 448-451, 2017.

[21] H. Yin and N. K. Jha, "A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles," IEEE Transactions on Multi-Scale Computing Systems, vol. 3, no. 4, pp. 228-241, 2017.

[22] B. D. Kanchan and M. M. Kishor, "Study of machine learning algorithms for special disease prediction using principal of component analysis," in 2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC), pp. 5-10, 2016.

[23] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh, and V. Mareeswari, "Prediction of cardiovascular disease using machine learning algorithms," in 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1-7, 2018.

[24] L. K. Tolentino, R. M. Aragon, W. R. Tibayan, A. Alvisor, P. G. Palisoc, and G. Terte, "Detection of Circulatory Diseases Through Fingernails Using Artificial Neural Network," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-4, pp. 181-188, 2018.

[25] R. J. Jorda, K. E. Bartolome, C. Buban, M. A. De Los Reyes, A. Maingat, M. J. Oledan, and L. K. Tolentino, "Ovulation hormonal imbalance recognition system using saliva's conductivity analysis based on potentiometric method," Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, no. 1, pp. 176-183, 2019.