# Machine learning applied to diabetes dataset using Quantum versus Classical computation

Danyal Maheshwari
eVida Research Group
University of Deusto
Bilbao, Spain
danyal.maheshwari@deusto.es

Begoña Garcia-Zapirain
eVida Research Group
University of Deusto
Bilbao, Spain
mbgarciazapi@deusto.es

Daniel Sierra-Soso
Computer science department
University of Louisville
Louisville,KY, USA
d.sierrasosa@louisville.edu

2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) | 978-1-6654-1589-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISSPIT51521.2020.9408944

***Abstract*—This paper presents a Quantum versus classical implemented of Machine learning (ML) algorithm applied to a diabetes dataset. Diabetes is a Sixth deadliest disease in the world and approximately 10 million new cases are registered every year worldwide. Using novel Quantum computing (QC) along with Quantum Machine Learning (QML) techniques in the healthcare system to improve and accelerate the computing of existing ML models that allows the different approach to understanding the complex patterns of the disease. The proposed system tackles a binary classification problem of patients with diabetes into two different classes: diabetes patients with acute diseases and diabetes patients without acute diseases. Our study compares classical and quantum algorithms, namely Decision Tree, Random Forest, Extreme Boosting Gradient and Adaboost, Qboost, Voting Model 1, Voting Model 2, Qboost Plus, New model 1 and New Model 2 along with an ensemble method which creates a strong classifier from a committee of weak classifiers. The results we achieved using the validation metrics of the New Model 1 showed an overall precision of 69%, a recall of 69%, an F1-Score of 69%, a specificity of 69% and an accuracy of 69% on our diabetes dataset, with an increase of the computation speed by ∼55 times in comparison of the classical system. Our study has proved that QC improves the computational speed and its inclusion in medical applications will deliver faster results to physicians and caregivers.***

***Index Terms*—Binary classification, Type 2 Diabetes Mellitus, Qboost ,Quantum vs Classical, Quantum computing.**

## I. Introduction

Diabetes is the sixth major cause of deaths in the globe, but Type 2 diabetes mellitus is a leading public health issue that presents a considerable rising pervasiveness [1][2]. 463 million people are living with diabetes, 232 million people are living in undiagnosed diabetes and 4.2 million mortalities happened in 2019 worldwide [3]. Diabetes can affect some of the vital organs of the human body, which may cause some of the major diseases such as blindness, renal failure, heart attack, limb amputation, and stroke. The increasing risk factor of diabetes projected that diabetic patients may rise to 700 million by 2045 worldwide, 374 million are growing risk of evolving Type 2 Diabetes, approximately 1.1 million children are living with diabetes and 20 million newborn babies are affected by diabetes during pregnancy [3]. To reduce the intricacy and improve the quality of T2DM patient's life, which maybe cause to increase the healthcare cost. Some research's shows that the controlling of the glycemic level is the most significant aspects to impede the organ detriment and other intricacies of T2DM. The decrease of glycemia in patients with diabetes substantially reduces the ailment and fatality rate [4][5]. Several studies have worked on analyzing ML and DL to predict acute diseases on the patient with T2DM [1][6][7][8]. All of them are classical approaches which uses traditional computers. This paper have a contribution of using the QC algorithms, Quantum processing unit (QPU) and compares the algorithm with the classical approach.

QC techniques are growing exponentially and numerous fields have achieved the extensive evolution. QC signifies that techniques are ready to attain the point where real-world applications must be enclosed in its outlook [7]. Can be QC along with QML techniques in the healthcare system to improve and accelerate the computing of the existing ML model that allows the different approaches to understand the complex patterns of the disease. The Quantum and ML algorithms such as Qboost classifier implement on the diabetes dataset, to compares the prediction and performance with a classical classifiers (Decision Tree, Random Forest, XGB and Adaboost). This aims to provide useful information for future research in the application of real-world datasets to QML techniques [9]. In this paper, we compare the Classical and Quantum algorithms with accuracies, precision, specificity, and computation time.

This paper is structured into 5 sections. Section I is about the introduction of QC, ML, and diabetes. Section II presents the patient's data and analysis of the variables of the dataset. Section III is all about the quantum and classical algorithms. Section IV covers the experiment design and elaboration of results whereas the conclusion is given in the last section.

## II. Materials

### A. Patient and Data

This research is based on the electronic health care records (EHRs) of Osakidetza (Basque Health Service). For this research, the analytical and clinical parameters information was acquired from the PREST Database (DB). In such DB, analysis coded according to the International Classification of Diseases (ICD-9-CM), and the encoding scheme applied for drugs is the anatomical, therapeutic, and chemical categorization system. For these purposes, in 2010, the PREST (the DB of the Stratification Program) was launched with the aim of categorizing the Basque citizens using the Johns Hopkins Adjusted Clinical Groups case-mix system [10]. A more comprehensive illustration is acquired from the preceding bibliography [11]. Despite the fact that type 2 diabetes can arise at any age but more often at age of 40 years or later, as in result, the patients are under age of 35 were eliminated. There was established and accomplished 12 months of successful periods (year 1: from 01–09–2007 to 31–08–2008; year 2: from 01–09–2008 to31–08–2009; year 3: from 01–09–2009 to 31–08–2010; year4: from 01–09–2010 to 31–08–2011). Throughout the time, a patient was deemed as offering T2DM if the sickness on a set date was before the determined end-point (i.e., foregoing 01–09–2007, or foregoing 01–09–2008 and so on), furthermore, public insurance was expected to be possessed by the patient at the start of the calendar year but it was not necessary to have it for the whole calendar year. Thereby, several of patients treated as a having T2DM was 116 295 in year 1; 123 991 in year 2; 130 554 in year 3, and 134 421 in year 4. The investigation of T2DM overall samples were 149 015 [1].

*1) Variables and Analysis:* At this stage research variables are considered like Gender, Age, Body Mass Index, Hemoglobin, LDL Cholesterol, Systolic and Diastolic blood pressure of the patients. It may be noted that the relationship of these parameters with T2DM was not reflected. To illustrate the prevalence of intricacy (minimum basic dataset) and the hospital records were used. Hospital admission because of acute myocardial infarction (MI), major amputation or avoidable hospitalizations also known as ambulatory care sensitive conditions (ACSC) was determined separately for each observation period. In order to review the acute diseases, a list of 52 health problems was developed and specific criteria were defined to consider that active disease during the period from 01/09/2010 to 31/08/2011 by adapting a methodology previously reported by other authors [1].

From such 52 listed health problems, diabetes mellitus was precluded. Further, this set of data was divided into two parts: a) related acute comorbidities, that according to other studies, embraces seven diseases such ischemic heart disease, renal failure, stroke, heart failure, peripheral neuropathy, foot ulcers, and diabetic retinopathy, and b) unrelated acute diseases that follow to the other listed 44 health problems. The raw DB was formed by 321 variables.

## III. Quantum Computing and Artificial Intelligence

### A. Quantum Computing

Quantum computing (QC) is a fusion of Quantum Mechanics and Quantum Information [12]. The Quantum Information is information that contains the state of QC. QC has the properties of Quantum Mechanics like Superposition, Entanglement and Tunneling [12]. Quantum Information algorithms are used within the ML and artificial Intelligence in state of Quantum system. QML [13] is a field that aims to comply quantum algorithms to perform ML tasks [14]. QML makes a revolution in speed and performance [15].

### B. Artificial Intelligence

The binary classification (Duda et al 2001) is the classifier which distinguishes entities into two categories based on classification rule [16]. DWave system QC has been thoroughly studied as a discrete optimization engine, which assumes any tasks which formulated as a quadratic unconstrained binary optimization (QUBO) [17][18].

Training a binary classifier with the quantum adiabatic algorithm (Naveen et al, 2008), which explained that binary classification using qboost ensemble method is measurable to the QC. The tasks were designed in such a way that rapid linear superposition the group of weak classifiers and QC to employ and modify the weights and process of learning that attempts to reduce the training error and number of weak classifiers [17][19].

### C. Artificial Intelligence Models

*1) Decision Tree Classifier:* A decision tree employs a tree structure to sort the data [14]. It utilizes non-leaf nodes to design the data to a variety of decision rules and leaf nodes to identify labels (+1, -1) for every datum [20][21]. It is necessary to formulate a decision tree by altering with the entropy. Although controlling the depth of a decision tree obliquely determine the sub-domains of the diabetes dataset. Whereas $c_n$ is the feature detector, $w_n$ is the weights and b are biasing to be optimized.

$$C(x) = sign\left(\sum_{i}^{N} w_n c_n(x)\right) \tag{1}$$

$$C(x) = sign(w_n * x + b) \tag{2}$$

*2) Forest Classifier :* Random forest (RF) is an ensemble method, it merges the various weak classifiers to fabricate one strong classifier. RF use to sort of decision tree as a weak classifier which separates the weights in the result of an enhanced model [22][23].

*3) Extreme Gradient Boosting Classifier :* Extreme Gradient Boosting (XGB) is an ensemble method, which combines several weak classifiers to fabricate one strong classifier [24][25]. It constructs the classifier in a stage-wise manner like any other boosting approaches and generalizes by enabling optimization of an arbitrary differentiable loss function [18].

*4) AdaBoost Classifier :* AdaBoost is an ensemble method in which a classifier is constructed in an iterative fashion [26][27]. In each repetition, it chooses the one weak classifier and re-learns to reduce the weighted error function.

$$L = \sum_{n=1}^{N} \exp\left\{-y_n \sum_{s=1}^{S} w_n c_n(x_n)\right\} \tag{3}$$

$$R(\omega) = \; ||w||_0 = \lambda \sum_{n}^{N} w_n^0 \tag{4}$$

$$w* = \arg min_w \left(L(\omega) + R(\omega)\right) \tag{5}$$

Whereas, L($\omega$) is the loss function, and R($\omega$) is the regularization and clinch that the labels are not to be complex. The simple selection of the regularization term based on the 0-norm,$||w||0$, which gives the number of non-zero weights. Where $\lambda$ controls the relative importance of the regularization[18][19].

### D. Quantum Computing Model

*1) QBoost Classifier :* QBoost is the ensemble learning method using Quantum Annealing (QA). In order to use the maximum power of DWave QA, the preparation of the objective function with reference to QUBO is required [15]. To perform that, we altered AdaBoost by substituting standard weighted error function with a QUBO [19].

$$w* = \arg\min_{w} \left( \sum_{s} \left( \frac{1}{N} \sum_{n}^{N} w_n c_n(x_s) - y_s \right)^2 \right) + \lambda ||w||_0 \tag{6}$$

Where, $C(x)$ is the strong classifier which assemble repetitively,in each repetition, chooses one weak classifier and relearns to reduce a weighted error function. Their weights are modified and renormalized to ensure that the sum of all weights is equal to 1.

The first part illustrates the distinguished within the weak classifier and the correct label. The second part illustrates the degree of weak classifiers to be used in the final classifier. $\lambda$ is a regularization variable that adapts the number of weak classifiers which concerns the total Hamiltonian [19].

Over there is a requirement to maximize this Hamiltonian by acknowledging the first part as a cost of the objective function and the second part as a constraint. Although reducing with QA allows us to acquire an amalgamation of weak classifiers that best fit of the training data.

## IV. Experiment Design and Results

### A. Experiment Design

The training of the models is implemented using Python3 and ML with scikit learn library. The help of the API (Application Programming Interface) to attempt and acquire the D-Wave system Quantum Processing Unit (QPU). Whereas D-Wave system QC allows Python notebook to run on a dedicated QPU. The training data is used to build each classifier, and the test data is used to compare the classifiers predicted labels with given known test labels. The labels +1 and -1 are representing diabetes with acute diseases and diabetes without acute diseases, respectively. In order to create our training and testing data, we have split our whole data into 80% for training and 20% for testing, while keeping the same ratio of the classes in each subset (50% of +1 and 50% of -1).

Ensemble methods [25] construct a strong classifier to enhance the predictive model by gathering all the weak classifiers: like bagging, boosting, and voting to improved prediction.

The ensemble methods to train the classifiers with training data sets of the diabetes dataset as shown in fig 1. In the begging, any factor has a similar probability to emerge in the dataset, in boosting data factors are weighted before they are collected in the dataset. An additional difference is that bagging is parallelizable, but boosting is essential to be running one after another.

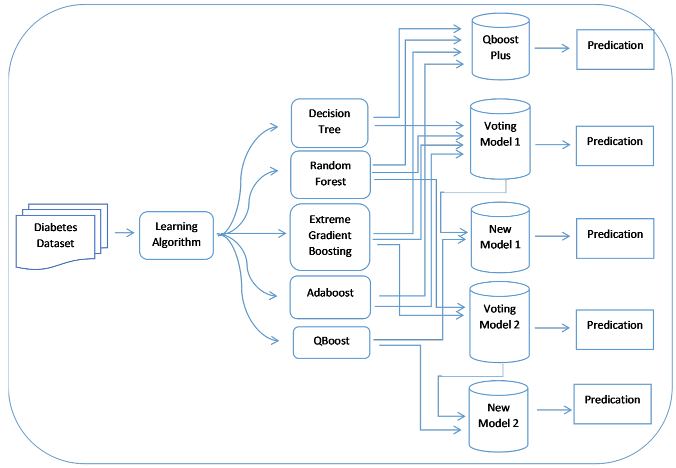


Fig. 1. Proposed Model

The Decision Tree, Random Forest, XGB, Adaboost and Qboost are used as a weak and strong classifier. In order to use the maximum power of DWave QA, the preparation of the objective function with reference to QUBO is required. Therefore, by replacing the exponential loss as in AdaBoost with the following quadratic loss where the regularization term is added to enable controlling of weight sparsity. QBoost uses Quantum Annealing to optimize the best combination of learners for the diabetes training dataset.

In the ensemble method, all classifiers are assembled repetitively. In each repetition, it chooses the one weak classifier and re-learns to reduces a weighted error function. The Voting model classifier has two typical requirements for its collection of weak classifiers: that there are several classifiers, and they should be diverse [28]. The voting model 1 classifier (Decision Tree, Random Forest, XGB and Adaboost) and voting model 2 (Random Forest and XGB) combine and create an individual voting system based on estimators and weights, to get an improved predication from the classical computer.

The QC classifiers like Qboost plus, New Model 1 and New Model 2 uses all classical classifiers. Qboost plus consists on (Decision Tree, Random Forest, XGB and Adaboost), New Model 1 consists on (Voting and Qboost), and similarly, New Model 2 comprises on (voting model 2 and Qboost) classifiers which build new classifier that runs on the Quantum system [19]. Selecting the right parameters to train the classifiers are considered to be the best approach to achieve better results. Ensemble learning involves preparing several weak predictors and combining the results of each of these predictors to obtain the final prediction result from Quantum state.

The maximum depth of a binary tree is the number of nodes along the longest path from the root node down to the farthest leaf node, tuning the weights, no of estimators and the hyperparameters of weak and strong classifiers along

with regularization of lambda ($\lambda$) for the Qboost, in terms of constructing the QUBO[18].
In order to analyze the behavior of the classifiers the following metrics are used: accuracy, precision, recall and specificity. Specificity is vital in medical disease classification, which gives additional information on the classifier of the disease. The obtained accuracies of each classifier are depicted in tables 1 to 10.
To analyze the implemented model of our system by fetching the well-known standards of the confusion matrix [29], like, precision, recall, accuracy, f1-score, and specificity [30][31]. The formulas for the individual entity are given below:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

$$Recall = \frac{TP}{TP + FP} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} or \frac{TP}{FN + TP} \tag{11}$$

Where TP stand for true positive, FP represents for false positive, TN means true negative, and FN depicts false negative.

*B. Results*

In this study, CPU-based algorithms (Decision Trees, Random Forest, XGB and Adaboost, Voting Model 1, and Voting Model 2) and QPU-based classifiers (Qboost, QBoostPlus, New Model 1 and New Model 2) are used to solve a binary classification problem
Table 1 to 10 are regrouped into table 11 using the validation metrics represented above. The classifiers on which we have focused are: Voting Model 1, Voting Model 2, Qboost plus, New Model 1 and New Model 2. Precisely our models accomplished better on diabetes dataset with an accuracy of 69.09%, 68.38%, 68.10%, 69.09% and 68.38% respectively.

TABLE I
DECISION TREE CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.66 | 0.68 | 0.66 | 0.6 |
| Diabetes without Acute Diseases | 0.63 | 0.6 | 0.63 | 0.68 |
| Weighted Avg | 0.65 | 0.64 | 0.63 | 0.64 |
| Accuracy | 0.6412 | | | |
| Time | 0.042s | | | |

TABLE II
RANDOM FOREST CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.65 | 0.69 | 0.67 | 0.69 |
| Diabetes without Acute Diseases | 0.67 | 0.64 | 0.66 | 0.64 |
| Weighted Avg | 0.66 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.6725 | | | |
| Time | 0.729s | | | |

TABLE III
XGBOOST CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.67 | 0.72 | 0.69 | 0.72 |
| Diabetes without Acute Diseases | 0.7 | 0.64 | 0.67 | 0.64 |
| Weighted Avg | 0.69 | 0.68 | 0.68 | 0.68 |
| Accuracy | 0.6820 | | | |
| Time | 0.809s | | | |

TABLE IV
ADABOOST CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.66 | 0.64 | 0.67 | 0.68 |
| Diabetes without Acute Diseases | 0.67 | 0.68 | 0.66 | 0.64 |
| Weighted Avg | 0.67 | 0.66 | 0.67 | 0.66 |
| Accuracy | 0.6607 | | | |
| Time | 3.367s | | | |

TABLE V
VOTING MODEL 1 CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.69 | 0.69 | 0.69 | 0.69 |
| Diabetes without Acute Diseases | 0.69 | 0.68 | 0.68 | 0.67 |
| Weighted Avg | 0.69 | 0.69 | 0.69 | 0.68 |
| Accuracy | 0.6909 | | | |
| Time | 19.202s | | | |

TABLE VI
VOTING MODEL 2 CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.67 | 0.72 | 0.70 | 0.72 |
| Diabetes without Acute Diseases | 0.70 | 0.64 | 0.67 | 0.64 |
| Weighted Avg | 0.68 | 0.68 | 0.68 | 0.68 |
| Accuracy | 0.6838 | | | |
| Time | 3.632s | | | |

Our purposed models have better accuracies, specificity's, and computational time in comparison with the classical Voting Model 1 and Voting Model 2. The Quantum algorithms perform better than the classical algorithms in terms of computation time of training and testing time of classical models, like Voting Model 1, Voting Model 2 are 19s, and 3.632s respectively, whereas the QPU running time of Qboost, Qboost plus, New Model 1 and New Model 2 classifiers 0.866s, 0.346s, 0.347s, and 0.347s respectively, but the quantum system is ~55 times faster than the classical system on diabetes data, as shown in table 11 and fig 2 which compares the results of all the classifiers. To ensure the authenticity and reliability of the

TABLE VII
QBOOST CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.66 | 0.61 | 0.63 | 0.6 |
| Diabetes without Acute Diseases | 0.64 | 0.69 | 0.66 | 0.7 |
| Weighted Avg | 0.65 | 0.65 | 0.65 | 0.65 |
| Accuracy | 0.6505 | | | |
| Time | 0.833s | | | |

TABLE VIII
QBOOST PLUS CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.67 | 0.64 | 0.69 | 0.65 |
| Diabetes without Acute Diseases | 0.69 | 0.71 | 0.66 | 0.71 |
| Weighted Avg | 0.68 | 0.68 | 0.68 | 0.68 |
| Accuracy | 0.6810 | | | |
| Time | 0.346s & 3.3s | | | |

TABLE IX
NEW MODEL 1 CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.69 | 0.69 | 0.69 | 0.69 |
| Diabetes without Acute Diseases | 0.69 | 0.68 | 0.69 | 0.69 |
| Weighted Avg | 0.69 | 0.69 | 0.69 | 0.69 |
| Accuracy | 0.6909 | | | |
| Time | 0.347s & 19.202s | | | |

TABLE X
NEW MODEL 2 CLASSIFIER

| Classes | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|
| Diabetes with Acute Diseases | 0.67 | 0.72 | 0.70 | 0.72 |
| Diabetes without Acute Diseases | 0.70 | 0.64 | 0.67 | 0.64 |
| Weighted Avg | 0.68 | 0.68 | 0.68 | 0.68 |
| Accuracy | 0.6838 | | | |
| Time | 0.347s & 3.632s | | | |

TABLE XI
COMPARISON OF CLASSICAL AND QUANTUM CLASSIFIERS

| Classifier | Accuracy | Prec-ision | Recall | F1-Score | Speci-ficity | Time |
|---|---|---|---|---|---|---|
| Decision Tree | 0.6412 | 0.65 | 0.64 | 0.65 | 0.64 | 0.042s |
| Random Forest | 0.6725 | 0.66 | 0.67 | 0.67 | 0.67 | 0.729s |
| XGB | 0.6820 | 0.69 | 0.68 | 0.68 | 0.68 | 0.809s |
| Adaboost | 0.6607 | 0.66 | 0.66 | 0.66 | 0.66 | 3.367s |
| Voting Model 1 | 0.6909 | 0.69 | 0.69 | 0.69 | 0.68 | 19.202s |
| Voting Model 2 | 0.6838 | 0.68 | 0.68 | 0.68 | 0.68 | 3.632s |
| Qboost | 0.6505 | 0.65 | 0.65 | 0.65 | 0.65 | 0.833s |
| Qboost Plus | 0.6810 | 0.68 | 0.68 | 0.68 | 0.68 | 0.346s 3.367s |
| New Model 1 | 0.6909 | 0.69 | 0.69 | 0.69 | 0.69 | 0.347s 19.202s |
| New Model 2 | 0.6838 | 0.68 | 0.68 | 0.68 | 0.68 | 0.347s 3.632s |

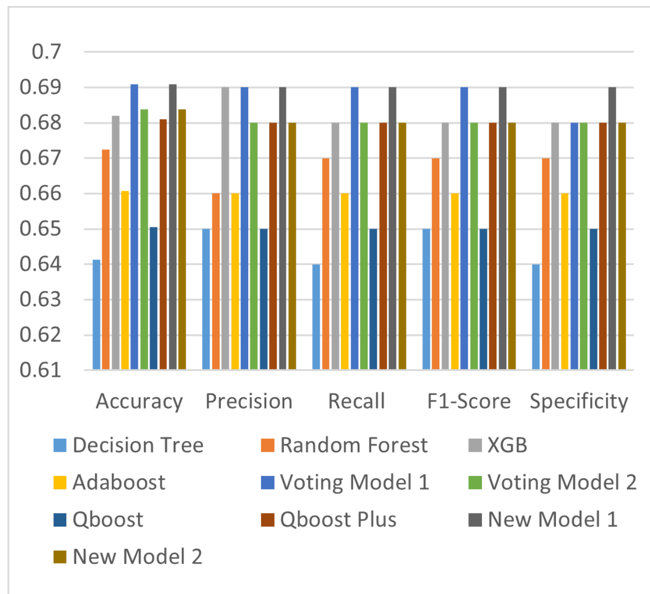


Fig. 2. Validation metrics for tested models

aforementioned classifiers in the medical field, a high specificity is required.

## V. CONCLUSION

In this study, we proved that our approach towards the binary classification of the diabetes dataset using the new models are feasible and faster, when they runs on a quantum system. The new models are based on the ensemble learning by combining classical and Quantum algorithms to construct a new classifier that runs on the Quantum system. The obtained results are satisfactory with the accuracies of 69% and 68.38%, a Recall and F1-score are 69% and 68.38% for the binary classification in the Classical and the Quantum system. The voting system is important for better accuracy, but the classical system takes bit longer time than the quantum to classify our diabetes data. We conclude that even including more and more algorithms in the Voting system results remains the same. The Voting Model 1 and Voting Model 2 have the almost same accuracy but Voting Model 2 takes less computational time in comparison of voting model 1. we choose the voting model 2 have better computation time than the voting Model 1. Quantum algorithms like New Model 1 and New Model 2 promises the similar accuracy's, recall, F1 scores and reduces the computational time. QC will be a useful and faster system in real-time application. The future direction is to be train dataset with different algorithms like qSVM and multi-label classification with of our improved system.

## ACKNOWLEDGMENT

This research has been funded by the Basque government IT-905-16 to eVIDA group and supported by Osakidetza that

dispense the database. The research codes were approved by the Clinical Research Ethics Committee of Euskadi (PI2014074), Spain. Informed consent was not obtained because patient health records were made anonymous and de-identified before analysis.

## REFERENCES


[1] D. Sierra-Sosa et al., "Scalable Healthcare Assessment for Diabetic Patients Using Deep Learning on Multiple GPUS," IEEE Trans. Ind. Informatics, vol. 15, no. 10, pp. 5682–5689, 2019.

[2] G. Du et al., "Metabolic Risk Factors of Type 2 Diabetes Mellitus and Correlated Glycemic Control/Complications: A Cross-Sectional Study between Rural and Urban Uygur Residents in Xinjiang Uygur Autonomous Region," PLoS One, vol. 11, no. 9, p. e0162611, Sep. 2016.

[3] International Diabetes Federation - Facts & figures. [Online]. Available: https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html. [Accessed: 15-Oct-2020].

[4] A. Ruiz-García et al., "Prevalence of diabetes mellitus in Spanish primary care setting and its association with cardiovascular risk factors and cardiovascular diseases. SIMETAP-DM study," Clin Investig Arter., vol. 32, no. 1, pp. 15–25, 2020.

[5] A. J. Scheen, "Cardiovascular Effects of New Oral Glucose-Lowering Agents DPP-4 and SGLT-2 Inhibitors," vol. 122, pp. 1439–1459, 2018.

[6] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, "Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Balanced Support Vector Machine," IEEE J. Biomed. Heal. Informatics, vol. 24, no. 1, pp. 235–246, 2020.

[7] K. Vidhya and R. Shanmugalakshmi, "Deep learning based big medical data analytic model for diabetes complication prediction," J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020.

[8] K.-M. Kuo, P. Talley, Y. Kao, and C. H. Huang, "A multi-class classification model for supporting the diagnosis of type II diabetes mellitus," PeerJ, vol. 8, p. e9920, 2020.

[9] D. Sierra-Sosa, J. Arcila-Moreno, B. Garcia-Zapirain, C. Castillo-Olea, and A. Elmaghraby, "Dementia Prediction Applying Variational Quantum Classifier," pp. 1–12, 2020 [arXiv:2007.08653].

[10] P. C. Austin, B. R. Shah, A. Newman, and G. M. Anderson, "Using the Johns Hopkins' Aggregated Diagnosis Groups (ADGs) to predict 1-year mortality in population-based cohorts of patients with diabetes in Ontario, Canada," Diabet. Med., vol. 29, no. 9, pp. 1134–1141, Sep. 2012.

[11] J. F. Orueta, A. García-Álvarez, M. García-Goñi, F. Paolucci, and R. Nuño-Solinís, "Prevalence and costs of multimorbidity by deprivation levels in the Basque Country: A population based study using health administrative databases," PLoS One, vol. 9, no. 2, Feb. 2014.

[12] C. Ciliberto et al., "Quantum machine learning: A classical perspective," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 474, no. 2209. Royal Society Publishing, 2018.

[13] M. Schuld and N. Killoran, "Quantum Machine Learning in Feature Hilbert Spaces," Phys. Rev. Lett., vol. 122, no. 4, p. 40504, 2019.

[14] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," Nature, vol. 549, no. 7671. Nature Publishing Group, pp. 195–202, 13-Sep-2017.

[15] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum deep learning," Quantum Inf. Comput., vol. 16, no. 7–8, pp. 541–587, 2016.

[16] G. Parmigiani, "Decision Theory: Bayesian," Int. Encycl. Soc. Behav. Sci., pp. 3327–3334, 2001.

[17] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, "Training a Binary Classifier with the Quantum Adiabatic Algorithm," pp. 1–11, 2008.

[18] R. Y. Li, R. Di Felice, R. Rohs, and D. A. Lidar, "Quantum annealing versus classical machine learning applied to a simplified computational biology problem," npj Quantum Inf., vol. 4, no. 1, 2018.

[19] H. Neven, V. S. Denchev, G. Rose, and W. G. MacReady, "QBoost: Large scale classifier training with adiabatic quantum optimization," J. Mach. Learn. Res., vol. 25, no. 2002, pp. 333–348, 2012.

[20] "1.10. Decision Trees — scikit-learn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/tree.html. [Accessed: 15-Oct-2020].

[21] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," IEEE Trans. Syst. Man Cybern., vol. 21, no. 3, pp. 660–674, 1991.

[22] "sklearn.ensemble.RandomForestClassifier scikitlearn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForest Classifier.html.

[23] M. Pal, "Random forest classifier for remote sensing classification," Int. J. Remote Sens., vol. 26, no. 1, pp. 217–222, 2005.

[24] "Introduction to Boosted Trees — xgboost 1.3.0-SNAPSHOT documentation." [Online]. Available: https://xgboost.readthedocs.io/en /latet/tutorials/model.html.

[25] "1.11. Ensemble methods — scikit-learn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html.

[26] "sklearn.ensemble.AdaBoostClassifier — scikit-learn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html.

[27] G. Rätsch, T. Onoda, and K. R. Müller, "Soft margins for AdaBoost," Mach. Learn., vol. 42, no. 3, pp. 287–320, 2001.

[28] "sklearn.ensemble.VotingClassifier — scikit-learn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html.

[29] S. Zahia, B. Garcia-Zapirain, I. Saralegui, and B. Fernandez-Ruanova, "Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging," Comput. Methods Programs Biomed., vol. 197, 2020.

[30] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. J. Aguirre, and A. M. Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," Sensors (Switzerland), vol. 20, no. 16, pp. 1–17, 2020.

[31] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, "Scikit-learn," GetMobile Mob. Comput. Commun., vol. 19, no. 1, pp. 29–33, 2015.