Hybrid Model of Quantum Transfer Learning to Classify Face Images with a COVID-19 Mask

Christian Soto-Paredes, Jose Sulla-Torres
Escuela de Ingeniería de Sistemas
Universidad Nacional de San Agustín de Arequipa
Arequipa, Perú

Abstract—The problem of the COVID-19 disease has determined that about 219 million people have contracted it, of which 4.55 million died. This importance has led to the implementation of security protocols to prevent the spread of this disease. One of the main protocols is to use protective masks that properly cover the nose and mouth. The objective of this paper was to classify images of faces using protective masks of COVID-19, in the classes identified as correct mask, incorrect mask, and no mask, with a Hybrid model of Quantum Transfer Learning. To do this, the method used has made it possible to gather a data set of 660 people of both sexes (man and woman), with ages ranging from 18 to 86 years old. The classic transfer learning model chosen was ResNet-18; the variational layers of the proposed model were built with the Basic Entangler Layers template for four qubits, and the optimization of the training was carried out with the Stochastic Gradient Descent with Nesterov Momentum. The main finding was the 99.05% accuracy in classifying the correct Protective Masks using the Pennylane quantum simulator in the tests performed. The conclusion reached is that the proposed hybrid model is an excellent option to detect the correct position of the protective mask for COVID-19.

Keywords—hybrid; quantum; classify; face; COVID-19; mask

I. INTRODUCTION

According to the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [1], which consolidates online information from the World Health Organization, the Chinese Center for Disease Control and Prevention, and the Johns Hopkins University as of September 2021, an estimated 219 million cases of people affected by COVID-19 [2] of which 4.55 million led to death. For this reason, security protocols have been implemented to prevent the spread of this disease. One of the main protocols consists of using protective masks that correctly cover the nose and mouth; even following safety protocols help reduce the incidence of other infectious diseases such as influenza, pneumonia, and mycobacterium tuberculosis [3]. Thus, this research contributes to compliance with the security protocol that establishes the use of protective masks in public places, the global impact is noted since it is not only a requirement demanded in all the cities of Peru, but it is also required in the rest of the countries around the world, in addition, reducing the number of people who dedicate themselves to the task of inspection and monitoring of compliance with this regulation has an impact on reducing costs and increasing the number of people verified. The literature presents Transfer Learning studies that try to solve this problem. A greater incidence in the use of ResNet-18 as a residual network is noted, with

significant results and accuracy greater than or equal to 90%, respectively.

Under this juncture, it was decided to classify the use that people give to masks in 3 classes: correct mask, incorrect mask, and no mask; it was decided to use Transfer Learning with ResNet-18 and enhance it with the use of quantum computing. An innovative hybrid model is born that represents an alternative for improvement to the classical methods of recognition, classification and/or prediction implemented in the literature. It is known that quantum computing helps in the simulation of quantum systems; it is in this subject that it is about simulating objects such as small molecules or macromolecules; quantum computing also supports the field of quantum optimization, helping to solve problems that have a cost that is intended to be minimized; also in stochastic physics, it helps to simulate random processes; It also supports risk analysis or simulation of probability distributions; even in cryptography; and finally in machine learning it allows machines to learn in a faster way since it allows representing multiple states at the same time. An additional motivation for proposing the hybrid model in question is that to date, there have been no Hybrid Quantum Transfer Learning investigations that seek to solve the problem of classifying face images using a COVID-19 protective mask.

A sample of 660 images in resolution of 1024×1024 pixels, of people facing the front wearing a protective mask, of both sexes (male and female), aged between 18 to 86 years, was considered, 68% of these images were used for training and the rest 32% for testing, for their use the images were transformed to $(3 \times 224 \times 224)$, in the training a resizing and cropping of the images was carried out, to improve the training cases; regarding the quantum simulator, Pennylane was used with four qubits, with an exact number of shots. Regarding the construction of quantum neural networks, these were made up of the embedding layer, the variational layers and the measurement layer; in this sense, 3 types of quantum variational layers were contemplated, those made with the Basic Entangler Layers template, the Strongly Entangling Layers template and a custom construction called Custom Layers, with Basic Entangler Layers being the one that provides the best results; for the optimization tasks of the training of the model as a whole, the performance was compared with the Adam optimizer and with the SGD with Nesterov Momentum, being SGD who provides the best results; to finalize an accuracy of 99.05% in the tests was obtained, which is satisfactory compared to the similar problems of Hybrid Quantum Transfer Learning reviewed in the literature, although it is true that

the accuracy, together with the precision, the recall and the f1-score are essential metrics to validate the relevance of the research, it should be noted that this study focuses on illustrating the innovative proposed method and the particular adaptation of the method by the authors, not on exceeding all existing methods in accuracy, it is in this sense that the development of the research and the techniques used are the most significant contribution left by this study.

After this introduction, this article is organized as follows: The works related to this article are explained in Section II the methodology used in this research in Section III, the experimental results in Section IV, discussions in Section V, conclusions in Section VI, and finally, future works in Section VII.

II. RELATED WORK

The bibliographic review contemplated various aspects such as Transfer Learning, ResNet, Hybrid Quantum Transfer Learning, and Datasets of images of faces with a mask.

A. Transfer Learning

Solving a problem involves gathering a large amount of information related to the solution of that problem; Transfer Learning is based on taking advantage of a large amount of data acquired and using it to solve another type of problem that shares certain characteristics. Thus, the author Yadav in [4] proposes a video surveillance system implemented in raspberry pi4, which is developed by Transfer Learning, with MobileNetV2 [5] as a pre-trained neural network and a single Shot Detector (SSD), which provides an 85% to 95% accuracy in detecting people using a protective mask.

There are Transfer Learning researchers that create their own datasets. This is how the authors Wang, Zhao, and Chen [6], propose a two-stage method that uses a Faster RCNN framework with Inception v2; your created dataset consists of 26,403 images of faces using masks; With their proposed model, they achieved 97.32% accuracy in simple scenes and 91.13% in complex scenes.

Transfer Learning modifies patterns already trained in specific characteristics and uses them to recognize similar patterns. Thus, using Inception v3 [7] in Face Mask Detection, the authors Jignesh Chowdary, Punn, Sonbhadra, and Agarwal [8], propose a model of Transfer Learning with the Simulated Masked Face Dataset (SMFD), achieving 99.9% during training and 100% during testing.

B. ResNet

ResNet [9] is known to allow deep training networks of more than 100 layers; In essence, it is a residual network which is based on the fact that some neurons connect with others in layers that are not necessarily contiguous, that is, jumping intermediate layers, considering this, the authors Loey, Manogaran, Taha and Khalifa [10] propose to use video object recognition to recognize people who use a protective mask, using for this purpose YOLO v2, here a feature extraction process is developed with ResNet-50 [9], achieving an accuracy of 81%.

The authors Addagarla, Kalyan Chakravarthi, and Anitha [11] proposed two models, FMY3 and FMNMobile; the latter uses Resnet SSD300 to tackle the task of recognizing people using masks, obtaining 98% accuracy and 99% recall rate.

The fundamental component of ResNet is the residual blocks; in a residual block, the input x is added directly to the output of the network by means of the jump connection; thus, the authors Jiang, Fan, and Yan [12] , propose a one-stage detector called RetinaFaceMask which implements a Feature Pyramid Network, achieving a prediction in the detection of people with a mask of 82.3% with MobileNet [5] and 93.4% with ResNet-18 [9].

To conclude, the authors Sethi, Kathuria, and Kaushik [13] present a two-stage detector that compares the ResNet-50, AlexNet, and MobileNet models, concluding that ResNet gives better results in the task of classifying faces with a mask in video surveillance, achieving an accuracy of 98.2%.

C. Hybrid Quantum Transfer Learning

Hybrid neural networks are made up of classical and quantum elements. There is the paradigm by which a pre-trained classical neural network is augmented with a variational quantum circuit [14]; this is how the paradigm called Hybrid Quantum Transfer Learning was born. Based on this approach is that the authors Mari, Bromley, Izaac, Schuld, and Killoran [15], propose to use this method for image recognition and quantum state classification; in their study, they are given the task of classifying images of bees and ants, among their The results obtained are 96.7% accuracy in the Pennylane simulator [16], 95% accuracy in IBM's ibmqx4 quantum hardware [17] and 0.80% in Rigetti's Aspen-4-4Q-A [18]. There are thus four ways to develop Transfer Learning, which are: classical to classical, classical to quantum, quantum to classical, and quantum to quantum; the classical to quantum approach is the one chosen for this research.

The authors Umer, Amin, Sharif, Anjum, Azam, and Shah [19], use Hybrid Quantum Transfer Learning with ResNet-18 and a 4-qubit quantum circuit to classify radiographic images of lungs of 3 types: COVID-19, Normal and Viral Pneumonia, with an accuracy of 99.7% in a quantum computer.

To run a Hybrid Quantum Transfer Learning model with ResNet-18, you can use simulators such as Pennylane [16], Qiskit-Aer [17], and Cirq [20]; you can also use real quantum hardware like the one provided by IBM: ibmq london, ibmq rome, among others, is how the authors Acar and Yilmaz [21] made their proposal in the devices mentioned above; They also used CT images of the lungs of 126 people with COVID and 100 ordinary people, achieving an accuracy of 90% in the simulator and 94% to 100% in quantum computers.

D. Dataset

The datasets contribute to the task of classifying images of people's faces using masks; so authors such as Cabani, Hammoudi, Benhabiles, and Melkemi [22] created a dataset that they called MaskedFace-Net, which consists of 137,016 images in resolution of 1024x1024 pixels, to which they added masks utilizing image editing [23], that is how they created two groups which they called:

- Correctly Masked Face Dataset (CMFD).

- Incorrectly Masked Face Dataset (IMFD).

To end, in [24], a dataset of more than 250,000 images of faces using masks is provided, it consists of high-quality images 1024x1024 pixels, these images were not edited by computer, so real people are shown as is using masks of 4 different ways, which can be interpreted as follows:

- Correct use of the mask, the one that covers the nose and mouth.

- Incorrect use, which does not cover the nose.

- Incorrect use, which does not cover the nose or mouth.

- Incorrect use, which does not have any mask.

III. METHODOLOGY

The objective of the study was to classify faces using protective masks from COVID-19 with a hybrid model of Quantum Transfer Learning; taking for this purpose a dataset of 660 people of both sexes (male and female), with ages between 18 to 86 years; in a diverse set of mask types and facial features. The classes identified for classification were 3: correct mask, incorrect mask, and no mask. The methodology used for the classification consists of 4 stages, elaborated under the approach of the Business Process Model and Notation (BPMN) [25] as shown in Fig. 1.
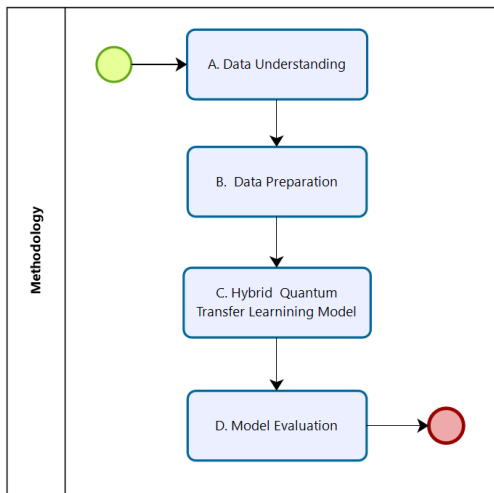

Fig. 1. Proposed Methodology for Classification (BPMN).

A. Data Understanding

The original dataset in [24] consists of more than 250,000 images divided into several parts; for the study, only part 2 of this dataset was taken, which consists of 40,000 images. This dataset is accompanied by a .csv file, which contains the detailed list of all the images; this is constituted as shown in Table I; this list is essential because it helped in the identification, understanding, and preparation tasks. Such as the subsequent processing of the sample images.

Explanation of the fields:

TABLE I. STRUCTURE OF THE DATASET LIST

Id	Type	User Id	Gender	Age	Name
Incremental	1-4	Numeric	M—F—N	Numeric	Text

- Type: Represents the four ways in which a person uses a mask, 1-correct mask, 2-incorrect mask (does not cover the nose), 3-incorrect mask (does not cover the nose or mouth), 4-incorrect mask (The person is not wearing the mask).

- User Id: Numerical value with which a person is identified.

- Gender: Male, female, or none.

- Age: Numerical value that is between 18 to 86 years old.

- Name: Sequence of alphanumeric characters with which the .jpg image is labeled.

The dataset organized by gender is shown in Table II, "None" corresponds to people whose gender was not registered; in addition, there is a predominance of people of the male gender 59.58%.

TABLE II. DATASET ORGANIZED BY GENDER

Gender	Number	Percentage (%)
Male	23832	59.58
Female	7012	17.53
None	9156	22.89

The dataset organized by age ranges is shown in Table III. From this, it is observed that the most significant population range is in the range of 18 to 30 years of age and that there are 1,948 people whose age was not registered, approximately 4.87%.

TABLE III. DATASET ORGANIZED BY AGE RANGES

Age	Number	Percentage (%)
18-30	27844	69.61
31-40	7732	19.33
41-50	2416	6.04
51-60	12	0.03
61-70	0	0.00
71-80	44	0.11
81-90	4	0.01
others	1948	4.87

B. Data Preparation

As a first step, the number of classes was reduced from 4 to 3, that is, to 30,000 images; since there is no significant difference between not bringing a mask and bringing a mask that does not cover the nose or mouth, the new output classes are:

- Correct Mask: This represents a correct use of the protective mask, which covers the nose and mouth (Fig. 2).

- Incorrect Mask: Representing an incorrect use of the protective mask, which does not cover the nose (Fig. 3).

- No Mask: This represents the absence of a mask that protects the individual (Fig. 4).

Fig. 2. Correct Mask.

Fig. 3. Incorrect Mask.

Second, a representative sample was taken, which is defined in (1), where:

n	Sample size.
N	Population size.
p	Probability of occurrence of the event studied.
q	Probability of non-occurrence of the event studied.
Z	Parameter that depends on the confidence level.
e	Estimation error.

$$n = \frac{N \times Z^2 \times p \times q}{e^2 \times (N-1) + Z^2 \times p \times q} \qquad (1)$$

The $30,000$ images considered became the initial population $N = 30,000$, in addition, a confidence level of 90% was estimated, that is, a $Z = 1,645$, an estimated error of 5.52%,
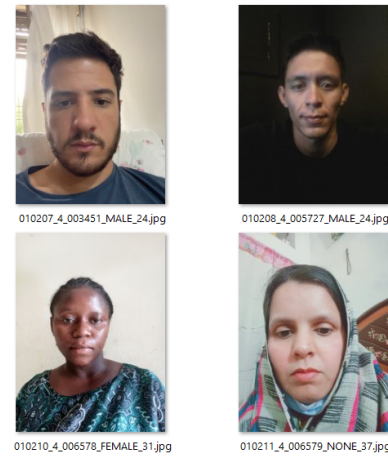

Fig. 4. No Mask.

and the values of p and q are calculated to have the same probability of occurrence 50% each. Executing the calculations with (1), it was obtained the selected sample was $n = 660$ in total, then, the next step was to randomly retrieve the images, considering 68% for training and 32% for testing as well as it's shown in Table IV.

TABLE IV. QUANTITIES FOR TRAINING AND TESTING

Phase	Percentage (%)	Total images	Total per class
training	68	450	150
testing	32	210	70

Then, transformations were made to the sample images; initially, these had a resolution of 1024×1024 pixels, but since it was the residual network ResNet-18 it required that they be transformed to the form $(3 \times 224 \times 224)$, it was also carry out a normalization with the mean and the standard deviation to $[0.485, 0.456, 0.406]$ and $[0.229, 0.224, 0.225]$ respectively, both for training and testing. Additional tasks such as random resizing and clipping were also performed on the training dataset, as well as a random horizontal rotation to improve the training cases. To finish, the data was loaded into a dictionary with the help of the Pytorch DataLoader [26], with a batch size of eight as shown in Fig. 5. and two workers to speed up the loading.

C. Hybrid Quantum Transfer Learning Model

Here the most critical topics in constructing the proposed hybrid model and its variants are detailed in 9 parts, which give rise to comparisons that trigger the final results.

1) Number of Qubits: It took four qubits, a qubit is the basic unit of quantum computers, just like a bit can take the value 0 or 1, however by the superposition principle, it could have a part of 0 and a part of 1, when measuring said qubit it will collapse towards one value or another; in definition, a qubit is a unit modulus vector in a complex two-dimensional vector space, the quantum states are $|0\rangle$ in (2) and $|1\rangle$ in (3).

Fig. 5. Transformations Made for Training in a Batch of Images of Size Eight.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad (2)$$

$$|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (3)$$

In addition, the state of a qubit can be represented by the "Bloch Sphere", this shows in the extreme north the state $|0\rangle$ and in the extreme south the $|1\rangle$ the rest of possible states would be represented along the surface of the sphere by $|\psi\rangle$, which is a linear combination of the states $|0\rangle$ and $|1\rangle$ as in (4), the state of a quibit is more likely to collapse towards $|0\rangle$ when it is further north and to collapse towards $|1\rangle$ when it is further south Fig. 6.
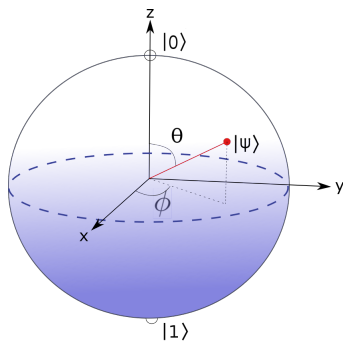

Fig. 6. Graphical Representation of a Qubit as a Bloch Sphere, shows the General States $|0\rangle$ and $|1\rangle$ and Arbitrary General States Represented by $|\psi\rangle$, where θ and ϕ are Real Numbers Such that $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$, any $|\psi\rangle$ can be Expressed as in (5).

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad (4)$$

$$|\psi\rangle = \cos(\frac{\theta}{2})|0\rangle + \epsilon^{i\phi}\sin(\frac{\theta}{2})|1\rangle \qquad (5)$$

2) Quantum Node and Differentiation Method: In pennylane [16], quantum computing is represented by objects composed of quantum nodes; a quantum node is used to declare a quantum circuit and relate it to a device that executes said circuit. Regarding the differentiation methods, for the simulation in pennylane [16], they are available: backprop,

adjoint, and reversible, and for real quantum computers: parameter-shift and finite-diff are available. For the present study, "reversible" was chosen because the quantum network will run in a simulator and give better results than the others.

3) Device and the Number of Shots: A "device" is the one that allows indicating in which hardware to execute the quantum circuit, for the given case "default.qubit" was chosen, the default local simulator of pennylane [16]; the number of "shots" was established as "exact," this is only possible because they are simulators, in real quantum computers the number of "shots" is determined by the equation proposed in (6), where:

c Confidence level.
Z Parameter that depends on the confidence level.
e Estimation error.

$$shots = (\frac{Z_{\frac{1-c}{2}}}{2 \times e})^2 \qquad (6)$$

For a confidence level of 95%, we have that $Z = -1.96$, and if an approximate error of 5% is taken, it is obtained that 384.16 is needed, which rounding would give rise to $shots = 384$.

4) Quantum Neural Network with Basic Entangler Layers: The proposed quantum neural network is similar to a classical fully connected neural network; it receives a vector of 512 features and has three outputs associated with the correct-mask and incorrect-mask and no-mask classes. The structure is composed of a preprocessing layer of type Linear, an activation function of type hyperbolic tangent, a layered architecture of Basic Entangler layers, and a post-processing layer. The layer mentioned above architecture is built with a variational quantum circuit or variational circuit [27]; this hybrid algorithm combines quantum and classical computing and is designed to be executed in quantum computers. To implement this type of variational circuit, you must first define a node with a quantum function that receives as parameters the number of features and the selected weights; we must resize the weights to the form: number of quantum layers x number of qubits. Then for each wire, you have to place a Hadamard gate (H) Fig.7, which operates on a single qubit, this is in charge of transforming the state $|0\rangle$ in $|+\rangle$ and the state $|1\rangle$ in $|-\rangle$, its matrix representation is observed in (7) as well as the representations of the $|+\rangle$ and $|-\rangle$ in (8) and (9) respectively.

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad (7)$$

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad (8)$$

$$|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \qquad (9)$$

Then three types of quantum layers are added, as detailed:

- Embedding layer: Here a layer consisting of single-qubit rotations was inserted, with rotation gates such

as RX (10), RY (11), and RZ (12), where ϕ is the angle of rotation, for this the template "Angle Embedding" was used, The graphic representation of this layer is shown in Fig. 7.

$$R_x(\phi) = \epsilon^{\frac{-i\phi\sigma_x}{2}} = \begin{bmatrix} \cos(\frac{\phi}{2}) & -i\sin(\frac{\phi}{2})) \\ -i\sin(\frac{\phi}{2})) & \cos(\frac{\phi}{2}) \end{bmatrix} \quad (10)$$

$$R_y(\phi) = \epsilon^{\frac{-i\phi\sigma_y}{2}} = \begin{bmatrix} \cos(\frac{\phi}{2}) & -i\sin(\frac{\phi}{2})) \\ \sin(\frac{\phi}{2})) & \cos(\frac{\phi}{2}) \end{bmatrix} \quad (11)$$

$$R_z(\phi) = \epsilon^{\frac{-i\phi\sigma_z}{2}} = \begin{bmatrix} \epsilon^{\frac{-i\phi}{2}} & 0 \\ 0 & \epsilon^{\frac{i\phi}{2}} \end{bmatrix} \quad (12)$$
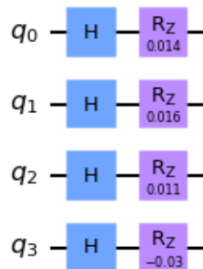

Fig. 7. Embedding Layer for 4 Qubits, with Hadamard Gates and Single Rotation Gates RZ (Graphic Generated with ibm Qiskit).

- Variational layers: Built with the Basic Entangler Layers template, which receives as parameters the weights, a vector of wires, and the axis on which to rotate, this is in charge of building the number of layers specified in the shape of the weights; This template inserts $CNOT$ gates (13) and single-qubit rotations (10), (11), and (12), as shown in Fig. 8.

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (13)$$
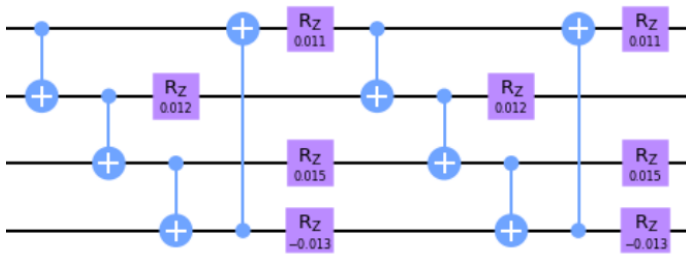

Fig. 8. Variational Layers with Basic Entangler Layers for four Qubits and a Depth of 2 Layers, with CNOT Gates and Single Quibit Rotations Gates RZ, as an Example, only Two Layers are Shown because, for the Evaluation of the Study, a Depth of 10 Variational Layers was Considered (Graphic Generated with ibm Qiskit).

- Measurement layer: This layer contains 4 Pauli Z operators (Fig. 9), which allow measuring the state of

the four qubits, which, when observed, will collapse towards one state or another; the matrix that represents it is shown in (14).

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (14)$$
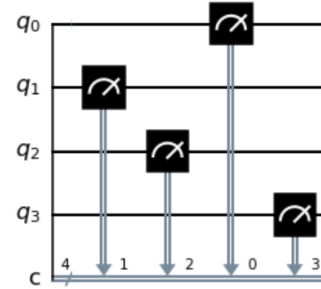

Fig. 9. Measurement Layer with 4 Pauli Z Operators (Graph Generated with ibm Qiskit).

5) Quantum Neural Network with Strongly Entangling Layers: In a similar way as in the previous section III-C4, the quantum neural network receives a vector of 512 features and provides three outputs corresponding to correct-mask, incorrect-mask, and no-mask, of Likewise, it consists of a processing layer, a post-processing layer and has a $tanh$ activation function, the difference lies in the architecture of the variational circuit that is built with the "Strongly Entangling Layers" template [28], so the embedding layer and the measurement layer are the same as seen in the previous section, Fig. 7 and Fig. 9 respectively, with the same operators seen in (10), (11), (12), (13), and (14).

The "Strongly Entangling Layers" template [28] is based on the circuit-centric classifier design, which is composed of single-qubit rotations interlaced with $CNOT$ gates; each single qubit G is a 2×2 unitary, as shown in (15).

$$G(\alpha, \beta, \gamma, \phi) = \epsilon^{i\phi} \begin{pmatrix} \epsilon^{i\beta}\cos\alpha & \epsilon^{i\gamma}\sin\alpha \\ -\epsilon^{-i\gamma}\sin\alpha & \epsilon^{-i\beta}\cos\alpha \end{pmatrix} \quad (15)$$

"Strongly Entangling Layers" does not support the "reversible" differentiation method, so pennylane [16] automatically selects the best possible method. In Fig. 10, this type of variational circuit is shown with a depth of 2 layers.

6) Quantum Neural Network with Custom Layers: Also in a similar way to section III-C4, the quantum neural network is completely connected as described in that section, the difference is that the architecture of the variational layers is based on the examples detailed in [15], this latest layer design proposal is based on interlacing of $CNOT$ gates and single-qubit rotations as seen in Fig. 11, the embedding layer is seen in Fig. 7 and the measurement layer in Fig. 9 respectively, also makes use of the operators seen in (10), (11), (12), (13) and (14).
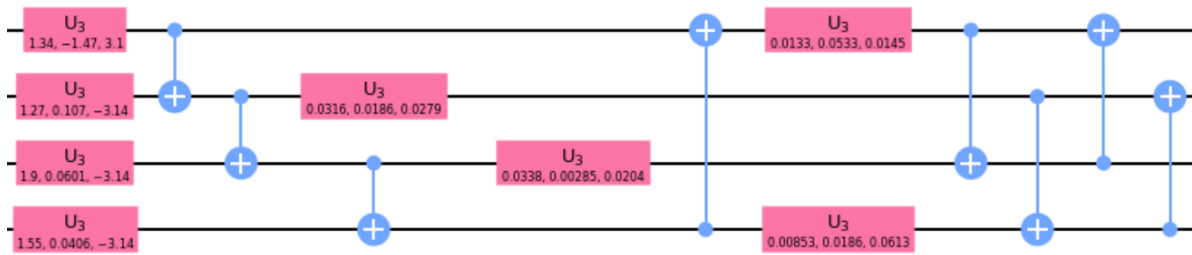
Fig. 10. Variational Layers with the use of the "Strongly Entangling Layers" Template for four Qubits and a Depth of 2 Layers, with $CNOT$ Gates and U_3 Rotations Gates that Represent Three Single-Qubit Rotations RZ, RY, and RZ Respectively, as an Example only Two are shown Layers because for the Evaluation of the Study a Depth of 10 Variational Layers was Considered (Graph Generated with ibm Qiskit).
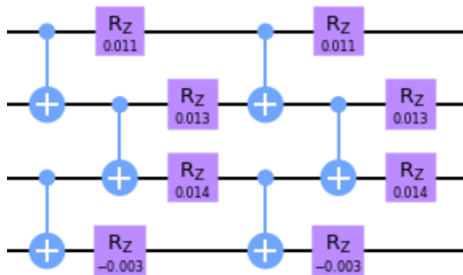

Fig. 11. Variational Layers with Basic Entangler Layers for four Qubits and a Depth of 2 Layers, with $CNOT$ Gates and Single Quibit Rotations Gates RZ, as an Example, only Two Layers are shown because, for the Evaluation of the Study, a Depth of 10 Variational Layers was Considered (Graphic Generated with ibm Qiskit).

TABLE V. RESNET-18 ARCHITECTURE

Layer Name	Filter Size	Stride	Padding	Number of Filters	Output Feature
Image Input	-	-	-	-	120x240x3
Conv1_x	7x7x3	2	3	64	60x120x64
	-	-	-	-	-
	3x3	2	1	-	30x60x64
Conv2_x	3x3x64	1	1	64	30x60x64
	3x3x64	1	1	64	
	3x3x64	1	1	64	
	3x3x64	1	1	64	
Conv3_x	3x3x64	2	1	128	15x30x128
	3x3x128	1	1	128	
	1x1x64	2	0	128	
	3x3x128	1	1	128	
	3x3x128	1	1	128	
Conv4_x	3x3x128	2	1	256	8x15x256
	3x3x256	1	1	256	
	1x1x128	2	0	256	
	3x3x256	1	1	256	
	3x3x256	1	1	256	
Conv5_x	3x3x256	2	1	512	4x8x512
	3x3x512	1	1	512	
	1x1x256	2	0	512	
	3x3x512	1	1	512	
	3x3x512	1	1	512	
	4x8	-	0	-	1x1x512

7) Hybrid Transfer Learning: About classic transfer learning, two types are considered: finetuning and feature extraction. In "finetuning," you start from a pre-trained model and update all the parameters for a new task; essentially, the entire model is completely re-trained. In "feature extraction," we start from a pre-trained model and only update the final layer's weights, with which we can now make predictions. Both transfer learning methods follow the following steps:

Step 1: Initialize the pre-trained model with the selected parameters.

Step 2: Reshape the final layer.

Step 3: Define an optimization algorithm.

Step 4: Carry out the training.

In addition to the above, it should be clarified that hybrid quantum computing receives its name for combining classical computing approaches with quantum computing. So to do hybrid quantum transfer learning, a classical pre-trained network must be taken, for this case, ResNet-18 (Table V), to which the final layer is reshaped to connect it with a quantum neural network whose architecture is based on a variational circuit that will be executed in a quantum computer.

In the present study, "feature extraction" was carried out, and the final layer of ResNet-18 is linked to the quantum neural network with 512 features in the manner described in section III-C4. Furthermore, let L be a layer, $n_0 \to n_1$ represents n_0 inputs and n_1 outputs, x an input vector and y an output vector as in (16).

$$L_{n_0 \to n_1} : x \to y \tag{16}$$

Then it is known that Q is a variational quantum circuit with E embedding layer and M measurement layer (17).

$$Q = M \circ Q' \circ E \tag{17}$$

Finally, to apply hybrid quantum transfer learning, the quantum neural network $\bar{Q}$ [15] has to be created as shown in (18), based on what is formulated in (16) and (17) .

$$\bar{Q} = L_{n_q \to n_{out}} \circ Q \circ L_{n_{in} \to n_q} \tag{18}$$

From all the above, it is possible to graphically observe the proposed hybrid quantum transfer learning model Fig. 12.

8) Optimization: The optimization process consists of the task of finding the best weights, minimizing the error in each iteration. For this, two methods were used: Adam [29] with a learning rate of 0.0004 and Stochastic Gradient Descent with Nesterov Momentum [30], the implementation of momentum
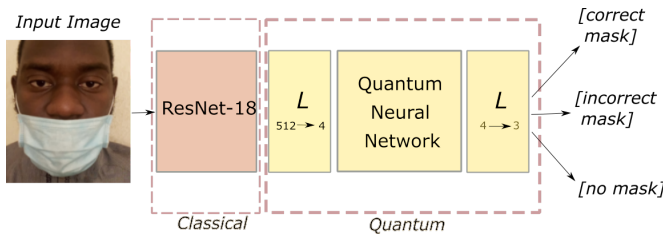
Fig. 12. The Proposed Hybrid Quantum Transfer Learning Model Receives an Image of a Person Wearing a COVID-19 Protective Mask, as Input and as Output a Classification (Correct Mask, Incorrect Mask, and no Mask) is Provided; on the Side of Classical Computing, it is Found ResNet-18 and on the Side of Quantum Computing is the Quantum Neural Network with 512 Features, four Qubits, and Three Outputs, this Quantum Neural Network is based on what is Formulated in (18).

Fig. 13. Composite Model Predictions with Strongly Entangling Layers for 30 Epochs, Optimized with SGD.

is described in (19) and (20), with a learning rate of 0.001 and a momentum of 0.9.

$$v_{t+1} = \mu * v_t + g_{t+1} \qquad (19)$$

$$p_{t+1} = p_t - lr * v_{t+1} \qquad (20)$$

Where p represents the parameters, g the gradient, v the velocity and μ the momentum.

9) Training: Is performed by first defining a training function that receives input parameters such as a model, a dictionary with the information of the loaded images, a loss function, in this case, "cross-validation", an Adam or SGD optimizer, several epochs to train, and an "is inception" flag which is used to prepare the function to work with Inception v3. It is worth mentioning that this training function was tested in transfer learning with the following classic models: resnet, alexnet, vgg, squeezenet, densenet, inception, and on the quantum side resnet-18 and densenet were implemented. This training function performs 2 phases: one of training itself and another that tests the time in question; with the optimizer, the best weights are chosen to allow the resulting model to classify between correct-mask, incorrect-mask, and no-mask. Finally, the results obtained by executing 10, 20, and 30 epochs were recovered, and each model was tested with the images intended for testing; a sample can be seen in Fig. 13.

D. Model Evaluation

To evaluate the model [31], the Precision (21), the Recall (22), and the F1-score (23) were used. For many classes C_i, fp_i represents false positive, tp_i true positive, fn_i false-negative, and tn_i true negative.

$$Precision = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \qquad (21)$$

$$Recall = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \qquad (22)$$

$$F1 - score = \frac{(\beta^2 + 1)(Precision)(Recall)}{(\beta^2)(Precision) + Recall} \qquad (23)$$

The repository where the source code is available, developed by the authors of this research, which allows to recreate the study and obtain the results presented, is available at [32].

IV. RESULTS

Tables VI and VII show the results of the classification, taking Accuracy and Training time as relevant factors; for 10, 20, and 30 epochs, considering a dataset of 660 images, 450 for training and 150 for testing, on the quantum side, there are four qubits, the number of shots was "exact," the depth of the variational layers was 10, The differentiation method was "reversible," and the processor used was the GPU provided by Google Colab [33], the selected simulator was "Pennylane" [16]. From now on, "Basic Entangler Layers" will be used interchangeably as "BEL," "Strongly Entangling Layers" as "SEL," and "Custom Layers" as "CL". The residual network selected for transfer learning was ResNet-18. From now on, all other results tables will work with the same hyperparameters previously described. The results of Tables VI and VII are divided by the type of optimizer used, "Stochastic Gradient Descent with Nesterov Momentum" and "Adam" respectively.

TABLE VI. RESULTS WITH THE STOCHASTIC GRADIENT DESCENT OPTIMIZER WITH NESTEROV MOMENTUM

Variational Layers	Epochs	Testing Accuracy (%)	Training Time (min)
BEL	10	97.14	48
BEL	20	98.57	97
BEL	**30**	**99.05**	**142**
SEL	10	95.71	105
SEL	**20**	**99.05**	**211**
SEL	30	98.57	309
CL	10	97.14	43
CL	**20**	**99.05**	**85**
CL	30	68.10	129

Tables VI and VII show that the best results were for "BEL" at 30 epochs, "SEL" at 20 epochs and "CL" at 20 epochs, the 3 with 99.05% accuracy and the 3 used the SGD as the optimization method. Once the three models with the best accuracy were identified, an analysis was performed based on the Precision, the Recall, and the F1-score to determine the

TABLE VII. RESULTS WITH THE ADAM OPTIMIZER

Variational Layers	Epochs	Testing Accuracy (%)	Training Time (min)
BEL	10	83.81	49
BEL	20	93.33	98
BEL	30	97.14	144
SEL	10	95.24	104
SEL	20	92.38	221
SEL	30	94.29	313
CL	10	87.62	42
CL	20	88.10	80
CL	30	98.57	127

model with the best classification. Tables VIII, IX, and X show the classification reports for the three best models mentioned above.

TABLE VIII. CLASSIFICATION REPORT FOR BASIC ENTANGLER LAYERS WITH STOCHASTIC GRADIENT DESCENT WITH NESTEROV MOMENTUM AT 30 EPOCHS

Class	Precision	Recall	F1-score
Correct Mask	0.96	0.99	**0.97**
Incorrect Mask	0.97	0.94	**0.96**
No Mask	0.99	0.99	**0.99**

TABLE IX. CLASSIFICATION REPORT FOR STRONGLY ENTANGLING LAYERS WITH STOCHASTIC GRADIENT DESCENT WITH NESTEROV MOMENTUM AT 20 EPOCHS

Class	Precision	Recall	F1-score
Correct Mask	0.91	0.87	0.89
Incorrect Mask	0.88	0.84	0.86
No Mask	0.89	0.97	0.93

TABLE X. CLASSIFICATION REPORT FOR CUSTOM LAYERS WITH STOCHASTIC GRADIENT DESCENT WITH NESTEROV MOMENTUM AT 20 EPOCHS

Class	Precision	Recall	F1-score
Correct Mask	0.84	1.00	0.92
Incorrect Mask	0.97	0.80	0.88
No Mask	0.99	0.97	0.98

Still, among the three, it is observed that the best result was for the model that uses variational layers built with the **"Basic Entangler Layers"** template and whose training optimizer was **"Stochastic Gradient Descent with Nesterov Momentum"** at **30 epochs**, given its **accuracy** of **99.05%** and its **F1-score** of **97% to classify "Correct Mask," 96% for "Incorrect Mask" and 99% to rate "No Mask"** respectively. Similarly, the confusion matrices Fig. 14, Fig. 15, and Fig. 16 confirm the analysis results mentioned above.

V. DISCUSSION

Table XI shows a comparison of image classification problems using Hybrid Quantum Transfer Learning models with ResNet-18 and four qubits; it can be seen that the proposed model shows the best accuracy with 99.05% when executed in a simulator, and also it is an encouraging result when compared
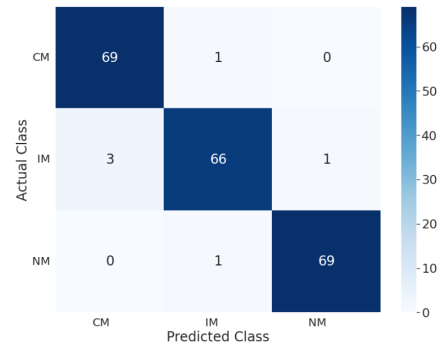

Fig. 14. Confusion Matrix for Basic Entangler Layers with Stochastic Gradient Descent with Nesterov Momentum at 30 Epochs.
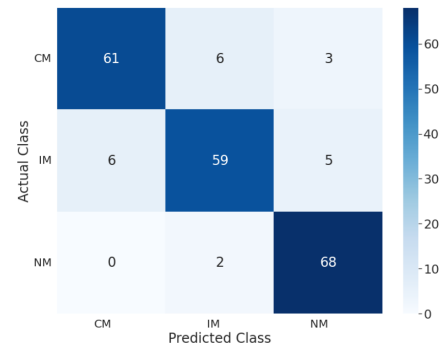

Fig. 15. Confusion Matrix for Strongly Entangling Layers with Stochastic Gradient Descent with Nesterov Momentum at 20 Epochs.

with the other results of real quantum computers. Although indeed, the accuracy, together with the precision, the recall, and the f1-score are essential metrics to validate the relevance of the research, it should be noted that the present study focuses on illustrating the innovative proposed method and the particular adaptation of the method on the part of the authors, not in surpassing in accuracy to all the existing methods, it is in this sense that the development of the investigation and the techniques used are the most significant contribution that the present study leaves.

TABLE XI. COMPARISON OF IMAGE CLASSIFICATION PROBLEMS USING HYBRID MODELS OF QUANTUM TRANSFER LEARNING WITH RESNET-18 AND 4 QUBITS

Autor	Classes	Images	Simulator Acc (%)	QC Acc (%)
[15]	2	ants, bees	96.7	95
[19]	3	lung covid-19	-	98
[21]	2	lung covid-19	90	94 - 100
Proposed	3	masked face covid-19	99.05	-

VI. CONCLUSION

A hybrid Quantum Transfer Learning model was developed with ResNet-18. Thanks to the inclusion in its quantum neural network of variational layers developed with the Basic Entangler Layers template and Stochastic Gradient Descent with Nesterov Momentum in its optimization, this is how it manages to enhance Its performance.
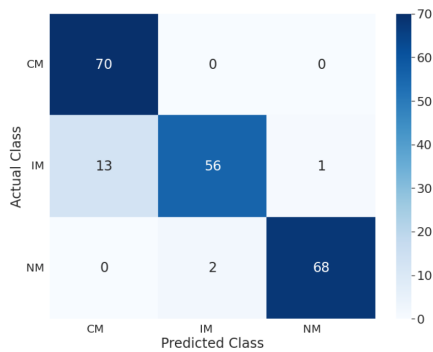
Fig. 16. Confusion Matrix for Custom Layers with Stochastic Gradient
Descent with Nesterov Momentum at 20 Epochs.

The evaluations carried out in classifying the use of a
protective mask (correct mask, incorrect mask, and no mask)
of images of 660 people of both sexes between 18 and 86
years old obtained an accuracy of 99.05%. These results
are satisfactory when compared with similar hybrid quantum
classification problems with ResNet-18 and four qubits in its
implementation.

VII. Future Work

The models trained with Strongly Entangling Layers at 20
epochs and Custom Layers at 20 epochs also obtained 99.05%
in accuracy but decreased in the levels of precision, recall and
f1-score, so it is proposed to improve the hyperparameters to
enhance their performance and future use.

From the consulted literature, it is known that there are other
quantum templates that could be more adaptable to the problem
of image classification, "CV Neural Net Layers" is one of
them, its research and the future proposal are recommended.

References

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1473309920301201

[2] (2021, Sep.) Covid-19 dashboard by the center for systems science and engineering at johns hopkins university. [Online]. Available: https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6

[3] O. Dadras, S. A. S. Alinaghi, A. Karimi, M. MohsseniPour, A. Barzegary, F. Vahedi, Z. Pashaei, P. Mirzapour, A. Fakhfouri, G. Zargari, S. Saeidi, H. Mojdeganlou, H. Badri, K. Qaderi, F. Behnezhad, and E. Mehraeen, "Effects of covid-19 prevention procedures on other common infections: a systematic review," *European Journal of Medical Research*, vol. 26, 2021.

[4] S. Yadav, "Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence," *International Journal for Research in Applied Science and Engineering Technology*, vol. 8, pp. 1368–1375, 07 2020.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[6] B. Wang, Y. Zhao, and C. L. P. Chen, "Hybrid transfer learning and broad learning system for wearing mask detection in the covid-19 era," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[8] G. Jignesh Chowdary, N. Punn, S. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of inceptionv3," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12581 LNCS, pp. 81–90, 2020, cited By 10. [Online]. Available: https://arxiv.org/abs/2009.08369

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[10] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, p. 102600, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670720308179

[11] S. Addagarla, G. Kalyan Chakravarthi, and P. Anitha, "Real time multi-scale facial mask detection and classification using deep transfer learning techniques," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 4402–4408, 2020, cited By 3. [Online]. Available: https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/en/covidwho-820057

[12] M. Jiang, X. Fan, and H. Yan, "Retinamask: A face mask detector," 2020.

[13] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of coronavirus spread," *Journal of Biomedical Informatics*, vol. 120, 2021, cited By 0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046421001775

[14] S. Adhikary, S. Dangwal, and D. Bhowmik, "Supervised learning with a quantum classifier using multi-level systems," *Quantum Information Processing*, vol. 19, no. 3, Jan 2020. [Online]. Available: http://dx.doi.org/10.1007/s11128-020-2587-9

[15] A. Mari, T. R. Bromley, J. Izaac, M. Schuld, and N. Killoran, "Transfer learning in hybrid classical-quantum neural networks," *Quantum*, vol. 4, p. 340, Oct 2020. [Online]. Available: http://dx.doi.org/10.22331/q-2020-10-09-340

[16] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," 2020.

[17] (2021, Sep.) Ibm quantum computing. [Online]. Available: https://quantum-computing.ibm.com/

[18] (2021, Sep.) Quantum computing service — amazon braket. [Online]. Available: https://aws.amazon.com/braket/

[19] M. J. Umer, J. Amin, M. Sharif, M. A. Anjum, F. Azam, and J. H. Shah, "An integrated framework for covid-19 classification based on classical and quantum transfer learning from a chest radiograph," *Concurrency and Computation: Practice and Experience*, p. e6434, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6434

[20] (2021, Sep.) Cirq an open source framework for programming quantum computers. [Online]. Available: https://quantumai.google/cirq

[21] E. Acar and İ. Yilmaz, "Covid-19 detection on ibm quantum computer with classical-quantum transfer learning," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/11/10/2020.11.07.20227306

[22] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "Maskedface-net – a dataset of correctly/incorrectly masked face images in the context of covid-19," *Smart Health*, vol. 19, p. 100144, Mar 2021. [Online]. Available: http://dx.doi.org/10.1016/j.smhl.2020.100144

[23] (2021, Sep.) Maskedface-net. [Online]. Available: https://github.com/cabani/MaskedFace-Net

[24] R. Kucev. (2021, Sep.) 500 gb of images with people wearing masks. [Online]. Available: https://www.kaggle.com/tapakah68/medical-masks-part2/metadata

[25] G. Decker, R. Dijkman, M. Dumas, and L. García-Bañuelos, *The Business Process Modeling Notation*. Springer, 09 2010, pp. 347–368.

[26] (2021, Sep.) Pytorch: An open source machine learning framework. [Online]. Available: https://pytorch.org/

[27] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New Journal of Physics*, vol. 18, no. 2, Feb 2016. [Online]. Available: http://dx.doi.org/10.1088/1367-2630/18/2/023023

[28] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," *Physical Review A*, vol. 101, no. 3, Mar 2020. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.101.032308

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[30] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1139–1147. [Online]. Available: https://proceedings.mlr.press/v28/sutskever13.html

[31] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing Management*, vol. 45, no. 4, pp. 427–437, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457309000259

[32] (2021, Oct.) Hybrid model of quantum transfer learning to classify face images with a covid-19 mask. [Online]. Available: https://github.com/cjsotopa/Quantum

[33] (2021, Sep.) Google colaboratory. [Online]. Available: https://colab.research.google.com/