International Journal of Applied Mathematics
Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HYBRID QUANTUM-CLASSICAL GENERATIVE ADVERSARIAL NETWORK FOR FAKE FACE DETECTION

Shahad Ghazi Hunbuli[1*], Eaman Alharbi[2], Lama Alkhuzayem[3]

[1*]*Department of Computer Science king Abdulaziz University* Jeddah, Saudi Arabia
Shanbuli0001@stu.kau.edu.sa

[2]*Department of Computer Science king Abdulaziz University* Jeddah, Saudi Arabia
ealraddadi@kau.edu.sa

[3]*Department of Computer Science king Abdulaziz University* Jeddah, Saudi Arabia
lalkhuzayem@kau.edu.sa

Abstract

Fake face generation is a growing problem for media integrity and online security. Deepfake images are now easy to create, and classical models often fail to detect them reliably. In this work, we introduce a Hybrid Quantum–Classical Generative Adversarial Network (HQCGAN) designed for fake face detection. A quantum layer was incorporated into the discriminator to enhance its capacity to distinguish real from synthetic facial images. Both HQCGAN and a classical GAN baseline were trained under identical experimental settings on the CelebA-HQ dataset. HQCGAN reached a test accuracy of 94.88% and an ROC-AUC of 0.9963, outperforming the classical GAN, which achieved 88.51% and 0.9565, respectively. The quantum-enhanced discriminator learned richer and more discriminative facial features, reducing misclassifications and improving generalization. To test robustness, the model was further fine-tuned on a combined dataset of StyleGAN2-ADA and FFHQ images. The consistent performance across different generative sources suggests that quantum integration can meaningfully contribute to more reliable deepfake detection frameworks.

Index Terms—Hybrid Quantum-Classical GAN (HQCGAN), Fake Face Detection, Deepfake Detection, Quantum Discriminator, CelebA-HQ Dataset, StyleGAN2-ADA, FFHQ Dataset,

I. INTRODUCTION

Generative Adversarial Networks (GANs) have become one of the most important advances in artificial intelligence. They were first introduced by Goodfellow et al. [1], combining a generator that creates synthetic data and a discriminator that evaluates the realism of this data. The interaction between the two networks enables the production of images that often appear indistinguishable from genuine ones. Generative Adversarial Networks (GANs) have significantly advanced areas such as computer vision, medical image synthesis, and data augmentation. However, these same developments have also raised serious security and ethical concerns. The capability to generate highly realistic human faces, voices, and actions has enabled the rise of deepfakes synthetic media designed to deceive both humans and automated recognition systems [2]. The widespread availability of Pretrained architectures such as StyleGAN [3] and CycleGAN [4] allow anyone to generate convincing fake content. Deepfakes have been used to spread misinformation, impersonate individuals, and manipulate public opinion. As a result, detecting synthetic images has become a

Received: August 05, 2025

critical research problem. The challenge has become more complex due to continuous improvements in generative models that reduce visible artifacts and enhance photorealistic quality [5].

Convolutional Neural Networks (CNNs) remain the foun- dation of most current deepfake detection systems. CNNs can learn hierarchical visual patterns and are effective at identifying small inconsistencies that occur during fake image generation. A Conv2D-based detector trained on 140,000 real and fake images achieved 94.54% accuracy on the OpenForensics dataset [6]. Another study applied Support Vector Machines with Principal Component Analysis to distinguish synthetic faces, achieving 96.8% accuracy in fake face classification [7]. A deep learning-based matching method was also developed to analyze both local texture and global structural cues for detecting manipulated faces [8]. These studies demonstrate that CNN-based detectors remain powerful tools for recognizing subtle artifacts, irregular color blending, and texture distortions introduced by GAN-generated images.

However, CNN-based procedures are limited. This usually reduces the performance in detecting images produced by invisible architectures or alternative compression plans. Most of the models were overfitted to features of the dataset like JPEG artifacts or light conditions. Therefore, their extrapolation is poor. This problem has been observed many times in the experiments that tested CNN detectors on cross-GAN and cross-dataset studies. The images produced by a different deepfake generator are not always detected by models that were trained on that one. These shortcomings demonstrate the need for more robust, architecture agnostic detection strategies.

Another area of new research is to investigate the use of GAN discriminators as detectors themselves. Given that discriminators are adversarially trained to distinguish between real and fake data during GAN learning, [9] and [10] both articles demonstrated that fine-tuned GAN discriminators are lightweight, high accuracy detectors without requiring retraining from scratch. They find that their findings are consistent with the hypothesis that adversarial components represent stronger statistical information regarding data realism as compared to standard supervised CNNs. This reuse of discriminators is a point between generative and forensic modeling.

In addition to classical methods of detection, Quantum Machine Learning (QML) also provides a novel depth. The quantum circuits take advantage of superposition and entanglement and encode data into higher dimensional space. Such quantum embeddings are able to learn relationships that otherwise are difficult to learn with classical neural networks. In this paper, a Hybrid Quantum-Classical Generative Adversarial Network (HQCGAN) of fake face detection is introduced. The model also incorporates quantum layer within the discriminator and compares its performance to that of classical GAN-based detectors. The proposed framework is trained and tested on the CelebA-HQ data. It is aimed at evaluating whether the hybrid design will improve accuracy, generalization, and stability in training in deepfake detection problems.

The main contributions are:

1) A hybrid GAN model integrating a quantum discriminator to improve the ability to distinguish real from synthetic faces.

2) Comparative evaluation between classical GAN-based, and HQCGAN-based detection frameworks.

The remainder of this paper is organized as follows. Section II reviews related studies on deepfake detection and the evolution of hybrid architectures. Section III presents the proposed methodology, including the model design, dataset preparation, and experimental setup. Section IV discusses the experimental results and performance evaluation. Finally, Section V concludes the paper and outlines potential directions for future research.

II. Related Work

Deepfake detection has become one of the most challenging problems in modern computer vision because of the rapid improvement of generative models, especially Generative Adversarial Networks (GANs). Several studies have investigated the use of GAN components, transfer learning, and even quantum circuits to enhance fake media detection accuracy.

To enhance deepfake detection, some researchers [9] explored the use of GAN discriminators as independent detectors rather than relying on conventional CNN-based classifiers. Multiple GAN architectures were trained, and the discriminator module was extracted as a standalone classifier to identify manipulated videos. Their framework used MesoNet as a baseline and evaluated performance on datasets such as DFDC and FaceForensics++, showing that while discriminators can recognize fake samples produced by their own generators, they generalize poorly to other sources. This work emphasized the importance of dataset diversity and suggested that standalone GAN discriminators may require ensemble or fine-tuning techniques to achieve more robust results.

Another study [10] proposed a direct approach using the discriminator of a trained Vanilla GAN as a detection engine. After adversarial training, the discriminator was fine- tuned as a binary classifier to distinguish real and synthetic images. The model achieved 100% accuracy in controlled tests and demonstrated that adversarially trained discriminators can learn latent forensic features often overlooked by conventional networks. The authors also noted limitations in cross-GAN generalization and highlighted the need for further research to improve robustness across datasets.

A more practical GAN-based deepfake image detection system was introduced in [11], integrating both generation and detection stages within a single architecture. The approach focused on real-time image verification using GANs as the backbone for both fake sample generation and detection. The system achieved over 90% accuracy on benchmark datasets such as Celeb-DF and DFDC, showing strong potential for online content authentication.

Recent efforts have also examined how quantum computing could improve deepfake detection performance. Quantum systems, through properties such as superposition and entanglement, can represent complex data distributions more efficiently than classical models [12]. In one of the early works, a quantum transfer learning (QTL) framework was introduced [13], combining a pre-trained ResNet-18 with a four-qubit quantum layer to classify fake and real facial images. The quantum component improved feature extraction and achieved 96.1% accuracy on a real-world deepfake dataset generated using commercial software.

Building on this direction, a Variational Quantum-Circuit Enhanced GAN (QC-GAN) was developed [14] to merge classical and quantum components in generative modeling. The hybrid generator, implemented with MindSpore Quantum, contained a variational quantum circuit followed by a classical layer, while the discriminator remained classical. Compared to a traditional GAN, the QC-GAN achieved superior Fre´chet Inception Distance (FID) scores with fewer parameters and training iterations. Although primarily focused on image generation, the findings showed that combining quantum and classical networks improves training stability and output quality, supporting future applications in fake-media detection.

A more recent contribution [15] integrated Quantum Transfer Learning (QTL) with a Class-Attention Vision Transformer (CaiT) architecture to detect forged videos. The model, trained on the DFDC dataset, achieved 90% accuracy and an ROC- AUC of 0.94. The QTL component enhanced the model's ability to capture subtle spatiotemporal inconsistencies, while CaiT improved global feature extraction. This hybrid design demonstrated that quantum-inspired layers can complement attention-based transformers for robust detection in dynamic video scenarios.

Further research has also examined the robustness and generalizability of face synthesis detection methods trained on high quality datasets such as CelebA-HQ and FFHQ [16]. The study analyzed how detectors behave when exposed to images generated by ProGAN, StyleGAN, and StyleGAN2 while using diverse architectures, including fingerprint-based networks, transfer-learning CNNs, and feature separation models. The work emphasized cross-dataset evaluation and the influence of data perturbations such as noise, compression, and blur. It offered important insights into the limitations of purely classical convolutional approaches and highlighted the need for more adaptive architectures that maintain stable performance across generative sources.

Overall, the existing works on quantum-enhanced models demonstrate the obvious tendency toward the combination of quantum computing with classical deep learning to address the shortcomings of conventional designs. These articles showed that quantum-classical architectures enhance the accuracy of models and indicate that quantum circuits can be useful in complex visual tasks. The present research will be based on this orientation by applying a hybrid quantum-classical GAN framework that will enhance the accuracy of fake-face detection and test its use on real facial datasets.

III. Methodology

This study follows an experimental comparative design. It evaluates how a traditional Deep Convolutional Generative Adversarial Network (DCGAN) performs compared to a new Hybrid Quantum-Classical GAN (HQCGAN) in detecting fake faces. Both models use the same generator structure. However, the HQCGAN's discriminator includes a quantum layer built with the PennyLane framework. The goal is to see if adding quantum operations can improve how the discriminator learns and classifies facial features.

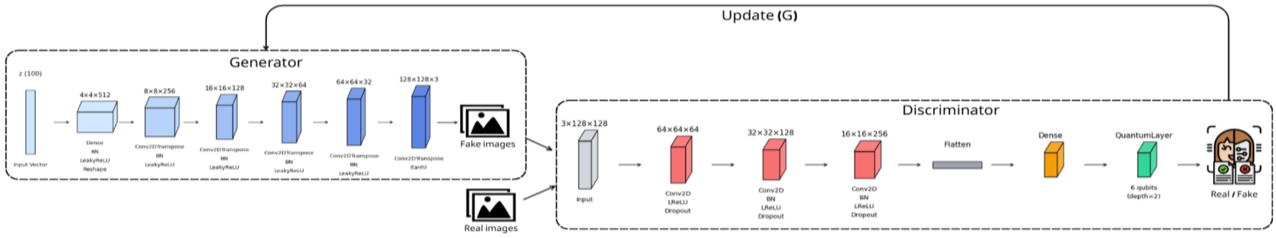


Fig. 1: Proposed Hybrid Quantum-Classical GAN (HQCGAN) Architecture

A. Dataset Description

The CelebA dataset [17] is a dataset of faces of famous people, it has been used as a sort of benchmark in computer vision studies. It has more than 200 thousand celebrity faces which are marked with 40 different attributes such as whether the individual is smiling, wearing glasses, or even turning their head slightly. Its variety is the thing that makes it useful. The photos are different in terms of lighting, pose and background which makes the models familiar with identifying faces in less controlled and more realistic situations. Nevertheless, other individuals have noted that since the data is based on celebrities, it is not necessarily reflective of the overall diversity of human faces that we come across in our daily lives.

CelebA-HQ [18] is based on that and is concerned with quality instead of quantity. It uses a subset of approximately 30,000 images chosen carefully and then processes it to eliminate artifacts, straighten the faces and enhances clarity using super-resolution. The original images are of high reso- lution of 1024x 1024 pixels. In my work, however, I used a resized 256×256 version from Hugging Face (*korexyz/celeba- hq-256x256*), which is a nice tradeoff. It retains the tiny details the skin texture, shades, the smooth lines at the edges of the features of the face and it is light enough to train effectively. The outcome is a dataset that is still sharp and realistic but does not overload the system, which is the best fit to determine the ability of a discriminator to distinguish real faces and generated ones.

The dataset was divided into 80% for training and 20% for testing. All images were further resized to 128×128 pixels and normalized to the range [-1, 1] to match the generator's output activation scale. The dataset was then converted into a TensorFlow data pipeline to enable efficient GPU-based loading and preprocessing during training. This diversity ensures that models trained on CelebA-HQ can generalize effectively across different face variations.

B. Model Architecture

The proposed HQCGAN architecture is illustrated in 1 follows a standard Generative Adversarial Network (GAN) design consisting of a generator and a discriminator trained in an adversarial manner. The generator learns to synthesize realistic facial images from random latent noise, while the discriminator learns to distinguish between real images from the dataset and fake images produced by the generator. Two discriminator configurations were implemented for compari- son: a classical DCGAN discriminator used as the baseline, and a hybrid quantum–classical discriminator that integrates the quantum layer described next. Both models share the same generator architecture and overall training setup to ensure a fair evaluation.

The generator adopts a classical DCGAN architecture composed entirely of transposed convolutional layers that progressively upsample the input latent vector into a (128*128*3) RGB image. Starting from a 100-dimensional Gaussian noise vector, the generator first projects and reshapes it into a (4*4*512) tensor. A sequence of five Conv2DTranspose layers follows, each accompanied by batch normalization and LeakyReLU activations. The number of filters decreases gradually (512, 256, 128, 64, 32), doubling the spatial resolution at each stage. The generator's final layer applies three filters with a tanh activation, producing images normalized to the range [-1, 1]. This setup helps maintain stable gradients and encourages realistic image synthesis.

For comparison, the baseline discriminator follows the standard DCGAN design as a fully classical benchmark. It includes several convolutional blocks that gradually reduce spatial resolution while capturing essential facial features. Each block uses a 4×4 convolution with stride two, followed by LeakyReLU activation and dropout for regularization. Batch normalization is applied to deeper layers to stabilize learning. After the final convolutional block, the features are flattened and passed through a dense layer before a sigmoid output neuron predicts the probability that an image is real. This baseline architecture is used to establish performance limits for a purely classical discriminator.

The hybrid discriminator, in contrast, extends this baseline by adding the QuantumLayer, which performs quantum feature transformation before the sigmoid output. After the convolu- tional feature extractor, the flattened feature vector is projected through a dense layer with tanh activation to scale its values within ([-1, 1]), matching the input domain of the quantum circuit. The final dense layer with sigmoid activation produce the real/fake probability. All other parts of the discriminator remain identical to the baseline version to isolate the effect of quantum integration.

Finally, the adversarial model combines the generator and discriminator into a unified HybridQuantumGAN class imple- mented in TensorFlow/Keras. This design ensures that both models: baseline DCGAN and hybrid HQCGAN are trained under identical conditions, enabling direct performance com- parison and clear attribution of improvements to the inclusion of the quantum layer.

C. Quantum Layer Design

The quantum layer in the proposed HQCGAN model was implemented using the PennyLane–TensorFlow interface to allow joint optimization with classical layers. The circuit was built on a default.qubit device configured with 6 qubits and a depth of 2 entangling layers. Classical features extracted by the convolutional part of the discriminator were first reduced through a dense layer with tanh activation function, ensuring their values matched the range suitable for quantum encoding. These features were then encoded into qubit states using AngleEmbedding, where each feature controls a rotation around the Y-axis. As shown in Figure 2 circuit consisted of parameterized RY gates applied to each qubit at every depth layer, followed by CNOT gates that connect neighboring qubits and an additional CNOT between the last and first qubit, forming a ring entanglement pattern. The design has interaction between features among qubits, at moderate computational cost. The circuit is used to measure the expectation value of the Pauli-Z operator of all qubits and the resulting output is a six-dimensional output vector. These expectation values are then passed through last classical layer to do final discrimination of real or fake images. The design is such that quantum and classical computations can be easily integrated with preservation of differentiability and numerical stability.

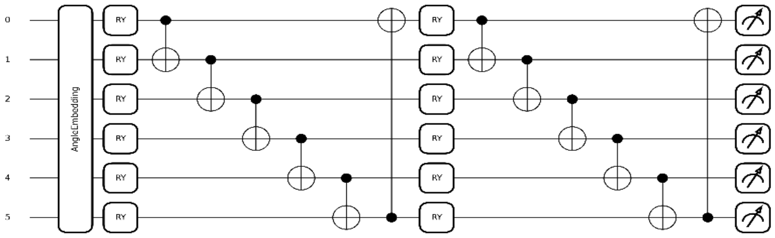


Fig. 2: Quantum Circuit Design

D. Implementation Details

Both models were trained on equal conditions so that there is fairness. Adam optimizer was applied with a learning rate of 0.0002 and β_1 = 0.5, a batch size of 128 and each model was trained with 50 epochs. The loss function of both the generator and discriminator were binary cross-entropy. The experiments were run on Python with Tensor Flow and Penny Lane and run on Google Colab with an NVIDIA A100. Random seeds were fixed for all frameworks to guarantee reproducibility.

TABLE I: Training Hyperparameters for Both Models

Parameter	Value
Optimizer	Adam
Learning rate	0.0002
β_1	0.5
Batch size	128
Epochs	50
Loss function	Binary Cross-Entropy
Frameworks	TensorFlow, PennyLane (Python)
Hardware	NVIDIA A100 GPU (Google Colab)
Random seed	42

IV. Results and Discussion

To evaluate the performance and effectiveness of the pro- posed HQCGAN, a series of experiments were conducted. The experiments were designed to measure the model's ability to distinguish between real and synthetic facial images and to assess its robustness compared to the classical GAN baseline. All models were trained and tested on the CelebA-HQ dataset following an 80/20 split. The evaluation metrics included Accuracy, ROC-AUC, Precision, Recall, and F1-score. The following subsections present the main experimental results and the additional fine-tuning analysis performed to further validate the generalization capability of the proposed hybrid model.

A. Experimental Results

To objectively compare the effectiveness of the proposed hybrid architecture with the classical GAN baseline, both models were trained and evaluated under identical experi- mental conditions. The HQCGAN achieved a test Accuracy of 94.88% and a ROC-AUC of 0.9963, outperforming the classical GAN, which achieved 88.51% Accuracy and 0.9565 ROC-AUC. Moreover, the HQCGAN demonstrated balanced performance across both real and synthetic face classes, with average Precision, Recall, and F1-score of 0.95. In contrast, the classical GAN showed lower and less stable performance (Precision = 0.90, Recall = 0.89, F1-score = 0.88), with a noticeable decrease in recall for real samples (0.80). These outcomes suggest that the integration of a quantum layer within the discriminator enhances the model's generalization ability and reduces misclassification errors compared with the baseline model.

TABLE II: Comparison of HQCGAN and Classical GAN performance on the CelebA-HQ dataset.

Model	Accuracy	ROC-AUC	Precision	Recall	F1-score
Classical GAN	0.8851	0.9565	0.90	0.89	0.88
HQCGAN (Proposed)	**0.9488**	**0.9963**	**0.95**	**0.95**	**0.95**

As shown in Table II, the HQCGAN consistently outperformed the classical GAN across all evaluation metrics. The quantum-enhanced discriminator achieved superior accuracy, robustness, and balance between precision and recall, indicating that quantum embedding and entanglement contributed to a richer feature representation of the facial data.

The superior results of the HQCGAN confirm the advantage of incorporating a quantum discriminator into the network architecture. The enhanced performance demonstrates that quantum feature encoding and entanglement enable more effective representation learning, allowing the discriminator to capture subtle differences between real and generated facial images. These improvements support the hypothesis that hybrid quantum-classical architectures can achieve higher discrimination accuracy and robustness compared with fully classical models.

The experimental findings demonstrate that the integration of a quantum layer within the discriminator significantly enhances the model's ability to distinguish between authentic and synthetic facial images. The proposed HQCGAN achieved superior performance compared to the classical GAN across all evaluation metrics, as reported in Table II. This improvement validates the hypothesis that quantum-enhanced feature encoding contributes to a more expressive and discriminative representation of facial features.

The approximately 6.4% increase in classification accuracy and 4% improvement in ROC-AUC highlight the advantage of hybrid quantum-classical architectures in capturing complex nonlinear patterns. The superior recall for fake samples (0.99) indicates that the HQCGAN is particularly effective at identifying subtle artifacts and texture inconsistencies introduced during generative synthesis. Meanwhile, the high precision for real samples (0.99) reflects the model's robustness in maintaining low false positive rates. Together, these results suggest that the quantum layer acts as a regularizer, improving the generalization ability of the discriminator and reducing overfitting.

From a representational perspective, the quantum circuit enables data embedding into a high-dimensional Hilbert space, where entanglement and superposition introduce richer correlations among features. This property allows the discriminator to establish clearer decision boundaries between real and generated faces. Such behavior aligns with theoretical pre- dictions that hybrid quantum-classical networks can achieve higher expressivity with fewer parameters compared to purely classical networks.

Another notable advantage observed during experimentation was the stability of the training process. The HQCGAN demonstrated smoother convergence and fewer discriminator-generator oscillations compared to the classical GAN. This stability is likely due to the stochastic quantum encoding, which injects controlled variability and prevents mode collapse, a common issue in classical GAN training.

Despite these strengths, the hybrid model introduces additional computational overhead due to quantum circuit simulation on classical hardware. Training time was moderately longer than that of the classical GAN, and scalability remains limited by the current simulation constraints. Nonetheless, the substantial improvement in detection performance justifies this cost, especially in forensic and security-sensitive applications where robustness is essential.

B. Additional Evaluation through Fine-Tuning on External Datasets

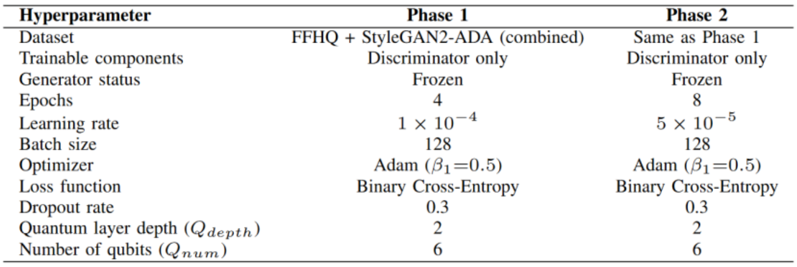
To further evaluate the generalization capability of the proposed HQCGAN, an additional fine-tuning experiment was conducted using datasets different from the original CelebA-HQ. Specifically, a subset of 10,000 real facial images was selected from the FFHQ-256 dataset hosted on Hugging Face [19], and an equivalent number of 10,000 synthetic images was generated using StyleGAN2-ADA from the official NVIDIA repository [20]. The combined dataset of 20,000 images was split into 80% for training and 20% for testing. This experi- ment aimed to assess the model's ability to generalize across different GAN architectures and real datasets while preserving strong discriminative performance.

The fine-tuning process was carried out in two consecutive phases, both using the same combined dataset but with progressively reduced learning rates to ensure smooth and stable adaptation.

- Phase 1 involved 4 epochs with a learning rate of 1×10^{-4}, allowing the quantum-enhanced discriminator to adjust to the mixed features of StyleGAN and FFHQ while preserving previously learned representations.

- Phase 2 continued for 8 additional epochs with a reduced learning rate of 5×10^{-5}, promoting gradual convergence and minimizing overfitting.

TABLE III: Fine-tuning hyperparameters for both phases.



Hyperparameter	Phase 1	Phase 2
Dataset	FFHQ + StyleGAN2-ADA (combined)	Same as Phase 1
Trainable components	Discriminator only	Discriminator only
Generator status	Frozen	Frozen
Epochs	4	8
Learning rate	1×10^{-4}	5×10^{-5}
Batch size	128	128
Optimizer	Adam (β_1=0.5)	Adam (β_1=0.5)
Loss function	Binary Cross-Entropy	Binary Cross-Entropy
Dropout rate	0.3	0.3
Quantum layer depth (Q_{depth})	2	2
Number of qubits (Q_{num})	6	6

Notes: Both phases share identical configurations except for the learning rate and number of epochs. Implementation was performed using TensorFlow and PennyLane on an NVIDIA A100 GPU.

The detailed configuration of both fine-tuning phases is summarized in Table III.

As shown in Table III, both phases maintained consistent configurations except for the gradual learning rate reduction, which enabled smoother convergence and improved discriminator stability. During both phases, the generator weights remained frozen, and only the quantum-enhanced discriminator was updated. This setup ensured that fine-tuning focused exclusively on improving the discriminator's ability to differentiate between real and generated facial images produced by various GAN architectures

The HQCGAN was found to be stable and high performing upon completion of 12 epochs. Table IV is the final evaluation metrics obtained with fine-tuning. These results show that quantum layer increases robustness and generalization across architecture, which allows HQCGAN to maintain high detection capabilities of both real and fake faces.

TABLE IV: Evaluation metrics of HQCGAN after fine-tuning (20% test split).

Metric	Value
Accuracy	0.9650
ROC-AUC	0.9935
Precision	0.9416
Recall	0.9915
F1-score	0.9659

In general, the findings affirm that’s because hybrid quantum-classical architectures can significantly benefit in the area of deepfake detection. By including a quantum layer into the discriminator, the accuracy and robustness were enhanced, proving how quantum-enhanced learning might be useful to detect subtle differences between real and fake facial images. This highlight the effectiveness of combining quantum compu- tation principles with deep learning models to advance reliable and generalizable detection frameworks.

V. Conclusion and Future Work

In summary, this paper proposed a hybrid quantum-classical GAN, which is dubbed as HQCGAN and developed to enhance the detection of fake face images. The concept was easy yet calculated keep everything concerning the classical GAN the same, with the exception of add a quantum layer to the discriminator. The result of that was that it became possible to observe the actual effect of quantum integration without other variables intervening. The results were clear enough to be convincing. The quantum version was more accurate, more stable, and and better generalized to other datasets and GANs architectures. These findings highlight the effectiveness of combining quantum computing principles with deep learning to advance reliable deepfake detection systems.

Further research is going to be based on a number of directions. A significant direction includes optimization of quantum circuit depth and structure to get a more optimal tradeoff between expressivity and cost. The other direction will be to test the HQCGAN on actual quantum hardware to test its scalability in non simulated settings. Also, further studies will look at Explainable Quantum AI (XQAI) methods to understand the decision boundaries of quantum enhanced discriminators, which have more transparency and trust in hybrid quantum-classical models. Lastly,extension of the HQCGAN framework to fake voice or video to demonstrate the extent to which this solution is indeed adaptable to a variety of media manipulation.

Acknowledgment

All praise and gratitude are due to Allah, the Exalted and Glorious, for His continuous guidance and blessings. The authors acknowledge King Abdulaziz University (KAU), Jeddah, Saudi Arabia, for providing a supportive research environment. This work was funded by the KAU Endowment (WAQF) and the Deanship of Scientific Research (DSR), whose technical and financial support is gratefully appreciated.

References

[1] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.

[2] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[3] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, 2019.

[4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2223–2232, 2017.

[5] L. Verdoliva, “Media forensics and deepfakes: An overview,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910– 932, 2020.

[6] D. Samal, P. Agrawal, and V. Madaan, “Deepfake image detection & classification using conv2d neural networks,” in *Proceedings of the Workshop on Advances in Computational Intelligence (ACI’23) at ICAIDS*, (Hyderabad, India), pp. 113–120, CEUR Workshop Proceedings, 2023. OpenForensics: Large-Scale Dataset for Multi-Face Forgery Detection.

[7] H. S. A. Kareem and M. S. M. Altaei, “Detection of deep fake in face images based on machine learning,” *Al-Salam Journal for Engineering and Technology*, vol. 2, no. 2, pp. 1–12, 2023.

[8] M. Favorskaya and A. Yakimchuk, “Fake face image detection using deep learning-based local and global matching,” in *Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications (SibDATA 2021)*, (Krasnoyarsk, Russia), pp. 1–8, CEUR Workshop Proceedings, 2021. Reshetnev Siberian State University of Science and Technology.

[9] S. A. Aduwala, M. Arigala, S. Desai, H. J. Quan, and M. Eirinaki, “Deepfake detection using gan discriminators,” in *2021 IEEE Seventh International Conference on Big Data Computing Service and Appli- cations (BigDataService)*, (Oxford, United Kingdom), p. 69–77, IEEE, Aug. 2021.

[10] G. S. Bisht, P. Gavit, A. Godhamgaonkar, H. Poshiya, and J. S. Pawar, “Deepfake detection using gan discriminators: Implementation and result analysis,” vol. 12, no. 1.

[11] R. Sonawane, A. Mandlik, P. Harnawal, S. Gade, and A. D. Londhe, “Deep fake image detection using GAN,” *International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS)*, vol. 7, no. 1, pp. 421–425, 2025.

[12] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[13] B. Mishra and A. Samanta, “Quantum transfer learning approach for deepfake detection,” *Sparklinglight Transactions on Artificial Intelli- gence and Quantum Computing*, vol. 02, no. 01, p. 17–27, 2022.

[14] Z. Shu, Z. Wu, Y. Yan, Z. Zhou, and Z. Huang, “Variational quantum circuits enhanced

generative adversarial network," *Computational Intel- ligence and Neuroscience*, vol. 2024, pp. 1–15, 2024.

[15] B. E. Katı, E. U. Ku¨c¸u¨ksille, and G. Sarıman, "Enhancing deepfake detection through quantum transfer learning and class-attention vision transformer architecture," *Applied Sciences*, vol. 15, p. 525, Jan. 2025.

[16] J. Sabel and B. Johansson, "On the robustness and generalizability of face synthesis detection methods," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2327–2336, IEEE, 2021.

[17] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, 2015.

[18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," in *International Conference on Learning Representations (ICLR)*, 2018.

[19] Bitmind, "Ffhq-256 dataset on hugging face." https://huggingface.co/datasets/bitmind/ffhq-256, 2024. Accessed: 2025-10-21.

[20] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Stylegan2-ada-pytorch: Official nvidia implementation." https://github.com/NVlabs/stylegan2-ada-pytorch, 2020. Accessed: 2025-10-21.

Received: August 05, 2025