A Hybrid Quantum-Classical Approach for Event Classification Using CERN Open Data

Khushi Verma[1], Palanivel R[2]
[1]*School of Information Science, Presidency University, Bengaluru, India*
[2]*Assistant Professor, Presidency University, Bengaluru, India*

***Abstract*—This paper presents a hybrid quantum-classical framework for classifying CERN CMS Open Data events containing charged leptons and neutrinos. A balanced dataset of 200,000 events was formed by combining 100,000 Wmunu and 100,000 Wenu samples, from which 26 engineered kine- matic and detector-level features were constructed. Classical baselines based on XGBoost, LightGBM, and CatBoost achieved 100.0% accuracy, 100.0% precision, 100.0% recall, 100.0% F1-score, and 1.000 ROC-AUC on the held-out test set. The quantum branch used median imputation, standardization, PCA reduction to four components, angle encoding on a 4-qubit circuit, and a PennyLane quantum kernel with an SVM classifier. The best quantum configuration achieved 99.0% accuracy, 98.04% precision, 100.0% recall, 99.01% F1-score, and 1.000 ROC-AUC on a balanced 100-event subset, with 49 true negatives, 1 false positive, 0 false negatives, and 50 true positives. The results confirm that compact quantum-kernel learning is competitive, although classical ensembles remain superior on this highly separable benchmark.**

***Index Terms*—Quantum Machine Learning, CERN Open Data, Quantum Kernel, Event Classification, Support Vector Machine, Hybrid Learning**

I. INTRODUCTION

Quantum machine learning (QML) has become an important direction for classification, kernel learning, and representation learning under near-term quantum constraints [1–5]. In parallel, high-energy physics is a natural application area because collider events are structured, high-dimensional, and rich in nonlinear relationships. Recent 2026 studies show that QML is increasingly assessed through robustness, hardware awareness, and scientific usefulness rather than conceptual novelty alone [6, 7].

This paper studies a hybrid quantum-classical event-classification pipeline using CERN CMS outreach data. The task is binary classification between $W \to \mu\nu$ and $W \to e\nu$ decay channels, both of which contain a neutrino inferred through missing transverse energy. The underlying notebook combines three classical ensemble baselines with a compact 4-qubit quantum-kernel model. The setup is intentionally practical, relying on public tabular data, standard preprocessing, low-qubit simulation, and direct comparison with strong classical learners.

The main contributions are as follows. First, the work assembles a balanced 200,000-event benchmark from CMS outreach datasets. Second, it develops a hybrid pipeline combining feature engineering, PCA compression, quantum angle embedding, and SVM-based inference. Third, it reports classical and quantum behavior through metric tables, confusion analysis, and comparative study. Fourth, it reorganizes the project into a formal IJIRT-style paper with literature review, algorithm design, environmental setup, and ethical discussion.

II. REVIEW OF LITERATURE

Recent 2026 literature shows that QML research is increasingly centered on practical model design. Dowling et al. studied adversarial robustness in quantum classifiers and emphasized the need for reliability guarantees [1]. Jha et al. compared quantum feature maps and showed that encoding choice materially affects kernel quality and decision boundaries [2]. Bravo-Montes et al. demonstrated that hybrid architectures must be evaluated together with realistic quantum noise [3].

Kernel and hybrid methods remain especially relevant for classification. Pinheiro et al. investigated quantum kernel and HHL based support vector machines for multi-class learning [8]. Martinez-Sabiote et al. showed that quantum-inspired similarity measures can preserve separability even without deep variational optimization [9]. Franco et al. combined SHAP-based feature selection with quantum phase classification, highlighting the value of interpretable reduction before quantum inference [10]. Adermann et al. additionally showed that quantum error detection can strengthen variational QML reliability [4].

Application-oriented studies also support compact hybrid pipelines. Singh et al. benchmarked MedMNIST on real quantum hardware [7]. Mourya et al. extended contextual quantum neural networks to predictive tasks [11]. Palanivel and Muthulakshmi studied error mitigation in secure qutrit distribution, quantum transfer fractal priority replay, dynamic-memory priority replay for robotics, and MaDi-based quantum prioritized experience replay [12–15]. Balasubramanian et al. emphasized that scalable hybrid quantum-inspired learning is especially valuable when classical preprocessing is tightly integrated with compact inference blocks [16]. Kong et al. proposed CQNAS for quantum-hybrid image classification, while Donaire et al. and Verdone et al. applied hybrid QML to liver- and heart-disease classification [17–19]. In high-energy physics, Yang et al. showed that QML classifiers can be useful in new-physics searches when benchmarked carefully against strong baselines [6]. The present work follows this practical benchmark philosophy on a more accessible public-data task.

III. METHODOLOGY

A. System Architecture

The complete workflow is shown in Fig. 1. Two CMS outreach datasets, Wmunu.csv and Wenu.csv, each containing 100,000 events, were loaded and labeled as class 1 and class 0, respectively. The unified feature table combined shared kinematic variables with detector-specific variables. Additional engineered features such as $\Delta\varphi$, transverse mass m_T, $|\eta|$, p_T +MET, p_T −MET, p_T /MET, $\cos(\Delta\varphi)$, and $\sin(\Delta\varphi)$ were also constructed.

The pipeline contains data acquisition, preprocessing, feature engineering, dimensionality reduction, and model inference. Classical preprocessing used median imputation and z-score standardization. The dataset was split with an 80:20 stratified train-test protocol. Three classical ensemble baselines were trained on the full standardized feature set. For the quantum branch, the standardized data were reduced to four principal components, and a balanced subset of 200 training events and 100 test events was selected to keep kernel evaluation tractable.

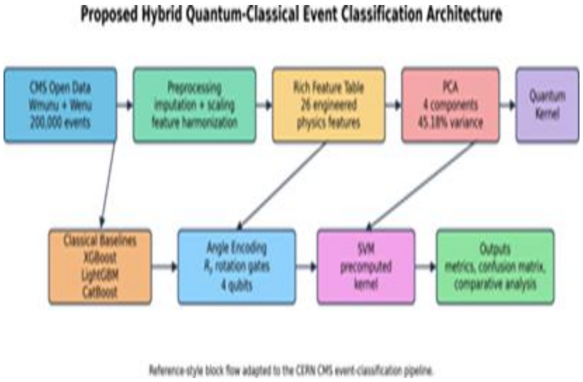


Fig. 1. System architecture of the proposed hybrid quantum-classical event-classification framework.

B. Model Setup

The quantum model uses a 4-qubit angle-encoding feature map followed by a precomputed-kernel SVM. Let $x \in \mathbb{R}^{26}$ denote the standardized engineered feature vector. PCA projects x into a reduced representation $z \in \mathbb{R}^4$:

$$z = W^T \hat{x}, \quad (1)$$

where $\hat{x}$ is the standardized input and W is the PCA loading matrix. The first four components retained a total variance ratio of 0.4518.

The reduced features were scaled to $-\frac{\pi}{2}, \frac{\pi}{2}$ and embedded using rotation gates:

$$\varphi(z) = \bigotimes_{i=1}^{4} R_y(z_i) 0^{\otimes 4}. \qquad (2)$$

Figure 2 shows the actual rotation-gate circuit extracted from the notebook output. In the broader notebook design, trainable $R_z(\theta_i)$ and final $R_y(\theta_{i+4})$ layers with nearest-neighbor CNOT entanglement were also considered to illustrate a hardware-efficient template.

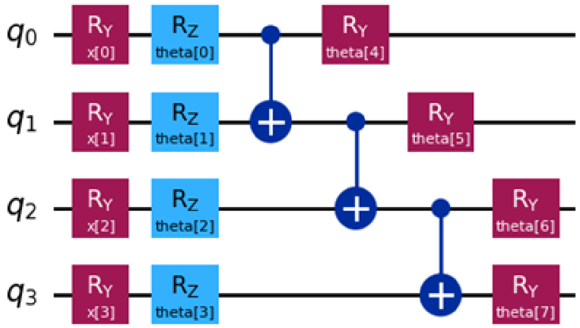


Fig. 2. Rotation-gate circuit used in the notebook for the 4-qubit hybrid quantum model.

The PennyLane kernel circuit computes the similarity between two reduced events:

$$K(\mathbf{z}, \mathbf{z}') = |\langle\varphi(\mathbf{z}')|\varphi(\mathbf{z})\rangle|^2 . \qquad (3)$$

The resulting kernel matrix is used by the SVM decision function

$$f(\mathbf{z}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{z}_i, \mathbf{z}) + b\right) . \qquad (4)$$

A sweep over $C \in \{0.1, 1, 5, 10, 20\}$ selected $C = 5$ as the best configuration.

IV. ALGORITHM DESIGN

Algorithm 1 summarizes the proposed event-classification workflow, and Fig. 3 provides the corresponding workflow diagram. Train-test arrays were also exported to HDF5 format for reproducibility [20].

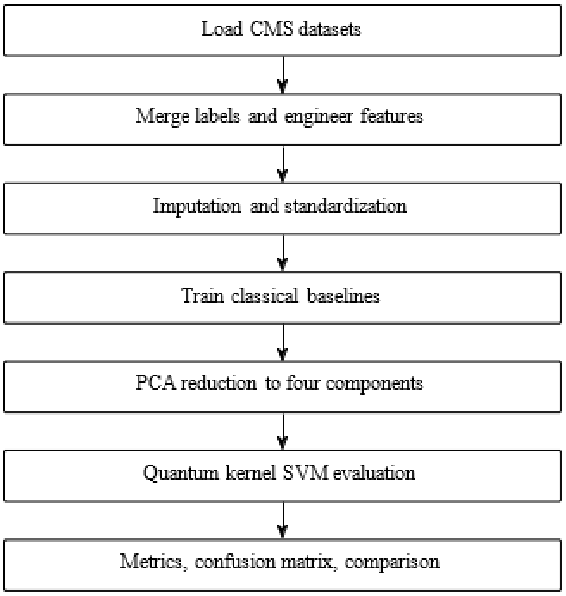


Fig. 3. Workflow diagram of the proposed hybrid quantum-classical algorithm.

```
Algorithm 1. Hybrid Quantum-Classical Event Classification
Require: CMS datasets $D_{\mu\nu}$ and $D_{e\nu}$
Ensure: Predicted class label $\hat{y}$
Load $D_{\mu\nu}$ and $D_{e\nu}$ from CERN Open Data
Assign labels 1 and 0 and merge into a unified dataset
Engineer kinematic and detector-level features
Apply median imputation and standardization
Split data into stratified train-test partitions
Train XGBoost, LightGBM, and CatBoost baselines
Apply PCA to obtain $z \in R^4$
Select balanced subset for quantum evaluation
Scale PCA outputs to $-\pi, \pi$
Encode features with 4-qubit $R_Y$ rotations
Compute $K_{train}$ and $K_{test}$
for each $C \in \{0.1, 1, 5, 10, 20\}$ do
    Train a precomputed-kernel SVM
    Evaluate accuracy, precision, recall, F1-score, and ROC- AUC
end for
Select the best quantum model and generate final predictions
```

V. ENVIRONMENTAL SETUP

The experiments were executed in a notebook-based Python environment using pandas, numpy, scikit-learn, xgboost, lightgbm, catboost, h5py, matplotlib, seaborn, Qiskit, and PennyLane. Quantum execution used PennyLane's default.qubit simulator, while Qiskit was used for circuit visualization. This software stack supports reproducible quantum simulation without requiring physical quantum hardware.

The dataset URLs are:
Wmunu.csv:
http://opendata.cern.ch/record/5205/files/Wmunu.csv
Wenu.csv:
http://opendata.cern.ch/record/5204/files/Wenu.csv

These files contain reconstructed event-level observables associated with charged leptons, missing trans- verse energy, and detector signatures. The processed train-test arrays were also exported to HDF5 format for reproducibility [20].

VI. RESULTS AND DISCUSSIONS

A. Model Performance

Table I reports the primary metrics. The three classical baselines achieved perfect separation on the rich feature space, while the quantum-kernel model achieved 99.0% accuracy, 98.04% precision, 100.0% recall, 99.01% F1-score, and 1.000 ROC-AUC on the balanced 100-event evaluation subset.

TABLE I Model Performance Summary

Model	Acc.	Prec.	Rec.	F1	ROC-AUC
XGBoost	1.000	1.000	1.000	1.000	1.000
LightGBM	1.000	1.000	1.000	1.000	1.000
CatBoost	1.000	1.000	1.000	1.000	1.000
Quantum Kernel SVM	0.990	0.980	1.000	0.990	1.000

For this project, loss is most appropriately discussed through log-loss for the classical probabilistic models rather than epoch-wise training curves, since boosted ensembles and precomputed-kernel SVMs are not optimized as standard epoch-driven deep networks. The very small classical log-loss values indicate extremely confident separation.

B. Predictions on Neutrino Events

The model does not observe neutrinos directly. Instead, neutrino-related information is inferred through missing transverse energy (MET), φ_{MET}, and transverse mass m_T . Both CMS datasets correspond to leptonic W -boson decays with a neutrino in the final state, so the task can be interpreted as classification of two neutrino-associated event topologies: muon-neutrino events and electron-neutrino events.

In practice, prediction is based on the combination of lepton kinematics and neutrino-sensitive observables. On the balanced quantum test subset, the model correctly identified all 50 class-1 events and 49 of 50 class-0 events. This indicates highly reliable prediction on neutrino-associated CMS events, although still slightly below the full classical baselines.

TABLE III Summary on Neutrino-Associated Events

Item	Observation
True negatives	49 class-0 events correctly predicted as class 0
False positives	1 class-0 event incorrectly predicted as class 1
False negatives	0 class-1 events incorrectly predicted as class 0
True positives	50 class-1 events correctly predicted as class 1
Interpretation	Strong separation after PCA compression and quantum embedding

These results indicate that the reduced quantum representation retains strong discriminative structure, although a small overlap remains after PCA compression and angle encoding.

This further supports the reliability of the quantum-kernel model for distinguishing the two neutrino-associated event classes under reduced feature representation. The single false-positive case also indicates that the remaining ambiguity is limited and does not significantly affect the model's ability to identify the target event topology.

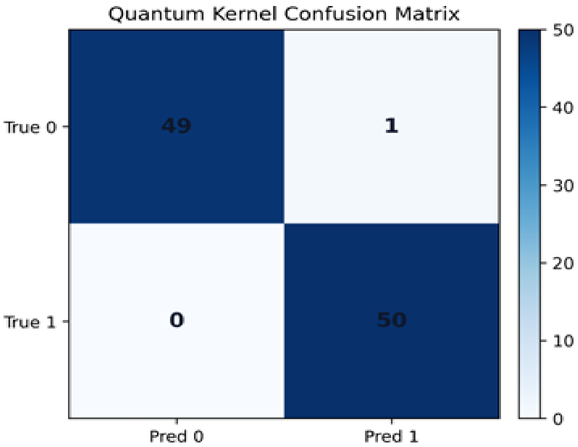


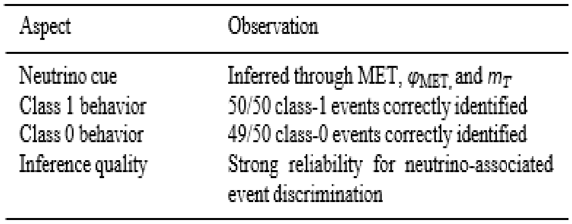
Fig. 4. Confusion matrix of the best-performing quantum-kern model

C. Predictions on Neutrino Events

The model does not observe neutrinos directly. Instead, neutrino-related information is inferred through missing transverse energy (MET), φMET, and trans- verse mass mT . Both CMS datasets correspond to leptonic W -boson decays with a neutrino in the final state, so the task can be interpreted as classification of two neutrino-associated event topologies: muon-neutrino events and electron-neutrino events.

In practice, prediction is based on the combination of lepton kinematics and neutrino-sensitive observables. On the balanced quantum test subset, the model correctly identified all 50 class-1 events and 49 of 50 class-0 events. This indicates highly reliable prediction on neutrino-associated CMS events, although still slightly below the full classical baselines.

TABLE III Prediction Summary on Neutrino-Associated Event



Aspect	Observation
Neutrino cue	Inferred through MET, φ_{MET}, and m_T
Class 1 behavior	50/50 class-1 events correctly identified
Class 0 behavior	49/50 class-0 events correctly identified
Inference quality	Strong reliability for neutrino-associated event discrimination

D. Comparative Analysis with Baseline Algorithms

The classical baselines outperformed the quantum branch by a small but real margin. XGBoost, LightGBM, and CatBoost each achieved 100.0% accuracy, whereas the quantum-kernel SVM reached 99.0%. This is expected because the rich feature set contains detector-specific variables that make the task strongly separable.

Nevertheless, the quantum model remains competitive. Its 4-qubit representation preserves most of the discriminative information after PCA compression, which is consistent with recent 2026 findings on hybrid quantum learning, feature-map sensitivity, hardware- aware classification, and application-oriented QML [7, 8, 17–19]. The main conclusion is therefore not quantum advantage, but quantum feasibility under severe dimensionality and resource constraints.

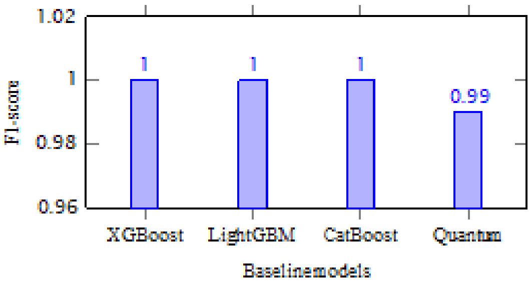


Fig. 5. Comparative analysis with baseline algorithms using F1-score.

E. Limitations and Ethical Considerations

Several limitations should be acknowledged. First, the classical models were trained on full stratified partitions, whereas the quantum model used a smaller balanced subset because kernel evaluation scales quadratically with sample size. Second, detector-specific variables make the benchmark easier than many realistic physics-analysis tasks. Third, the quantum model was evaluated in simulation rather than on noisy hardware, so the reported performance may overestimate real deployment behavior.

Ethically, QML studies should avoid overstating quantum superiority when benchmarks are small, curated, or unusually separable. Transparent reporting of preprocessing choices, subset sizes, and evaluation criteria is essential. Public scientific datasets should also be cited responsibly, with acknowledgment that outreach datasets simplify the complexity of full experimental workflows. The strongest claim supported here is that the proposed framework is a meaningful and reproducible hybrid QML demonstration rather than evidence of a general quantum advantage.

VII. CONCLUSIONS

This paper presented a revised hybrid quantum-classical event-classification study using CERN CMS Open Data and organized it into a formal IJIRT-style structure. The work combined 100,000 Wmunu events and 100,000 Wenu events into a balanced 200,000-event benchmark and engineered 26 kinematic and detector-level features relevant to leptonic W -boson decays with neutrinos. Three classical baselines, XGBoost, LightGBM, and CatBoost, each achieved 100.0% accuracy, 100.0% precision, 100.0 recall, 100.0% F1-score, and 1.000 ROC-AUC. The quantum branch reduced the feature space to four principal components, encoded the reduced representation through 4-qubit rotation gates, and used a PennyLane quantum kernel with a support vector machine. The best quantum configuration at $C = 5$ achieved 99.0% accuracy, 98.04% precision, 100.0% recall, 99.01% F1-score, and 1.000 ROC-AUC, with confusion matrix values of 49 true negatives, 1 false positive, 0 false negatives, and 50 true positives.

These findings show that low-qubit quantum-kernel classification can remain highly competitive even after aggressive dimensionality reduction. However, the study does not establish quantum advantage. The classical baselines are stronger on this benchmark, and the dataset is highly separable because it includes class-specific detector observables. The fairest interpretation is therefore that the proposed framework demonstrates the feasibility, interpretability, and reproducibility of hybrid QML for public particle-physics data. Future work should investigate more challenging event topologies, reduce dependence on highly revealing detector-specific fields, increase quantum sample size, include real hardware noise, and extend evaluation toward broader uncertainty-aware scientific inference.

REFERENCES

[1] N. Dowling, M. T. West, A. Southwell, A. C. Nakhl, M. Sevior, M. Usman, and K. Modi, “Adversarial robustness guarantees for quantum classifiers,” npj Quantum Information, vol. 12, no. 16, 2026. [Online]. Available: https://doi.org/10.1038/s41534-025-01129-3

[2] R. K. Jha, N. Kasabov, S. Bhattacharyya,Coyle, and G. Prasad, “Comparative performance analysis of quantum feature maps for quantum kernel-based machine learning,” Scientific Reports, vol. 16, no. 1, p. 8142, 2026. [Online]. Available: https://doi.org/10. 1038/s41598-026-39392-9

[3] J. A. Bravo-Montes, A. Martín-Toledano,Velasco-Gallego, G. Botella, A. A. del Barrio, and F. García-Herrero, “Design of a hybrid quantum machine learning architecture and analysis of quantum noise effects,” Scientific Reports, 2026. [Online]. Available: https://doi.org/10.1038/s41598-026-42216-5

[4] E. Adermann, H. Suzuki, and M. Usman, “Variational quantum machine learning with quantum error detection,” Quantum Machine Intelligence, vol. 8, p. 1, 2026. [Online]. Available: https://doi.org/10.1007/ s42484-026-00347-y

[5] L. Nadeem, A. M. Khokhar, M. F. Shaikh, F. H. Syed, and S. Zahid, “Quantum machine learning: A comprehensive survey of models, challenges, and future directions,” Theoretical Computer Science, p. 115923, 2026. [Online]. Available: https://doi.org/10.1016/j.tcs.2026.115923

[6] J.-C. Yang, S. Zhang, and C.-X. Yue, “A quantum machine learning classifier to search for new physics,” Journal of High Energy Physics, vol. 2026, no. 1, p. 23, 2026. [Online]. Available: https://doi.org/10.1007/JHEP01(2026)023

[7] G. Singh, H. Jin, and K. M. M. Jr., “Benchmarking medmnist dataset on real quantum hardware,” Scientific Reports, vol. 16,p.9017,2026. [Online]. Available: https: //doi.org/10.1038/s41598-026-35605-3

[8] G. Pinheiro, D. M. Slabbert, L. Kowada et al., “Quantum kernel and hhl-based support vector machines for multi-class classification,” EPJ Quantum Technology, 2026. [Online]. Available: https://doi.org/10.1140/ epjqt/s40507-026-00484-0

[9] A. Martínez-Sabiote, M. Skotiniotis, J. J. Bermejo-Vega, D. Manzano, and C. Cano, “Entanglement detection with quantum-inspired kernels and svms,” The Journal of Supercomputing, vol. 82, p. 100, 2026. [Online]. Available: https://doi.org/10.1007/ s11227-026-08229-7

[10] G. S. Franco, F. Mahlow, P. M. Prado et al., “Quantum phases classification using quantum machine learning with shap-driven feature selection,” Quantum Information Processing, vol. 25, p. 106, 2026. [Online]. Available: https://doi.org/10.1007/s11128-026-05132-1

[11] S. Mourya, H. Leipold, and B. Adhikari, “Contextual quantum neural networks for stock price prediction,” Scientific Reports, vol. 16, p. 4454, 2026. [Online]. Available: https://doi.org/10.1038/s41598-025-34413-5

[12] R. Palanivel and P. Muthulakshmi, “Error mitigation using quantum neural q network in secure qutrit distribution on cleve’s protocol on quantum computing,” Quantum Information Processing, vol. 23, p. 147, 2024. [Online]. Available: https://doi.org/10.1007/ s11128-024-04342-9

[13] “Design and analysis of quantum transfer fractal priority replay and mirdad priority loss algorithms for quantum reinforcement learning,” in Data Management, Analytics and Innovation, ser. Lecture Notes in Networks and Systems. Singapore: Springer, 2024, vol. 997, iCDMAI 2024. [Online]. Available: https://doi.org/10. 1007/978-981-97-3242-5_27

[14] “Design and analysis of parallel quantum transfer fractal priority replay with dynamic memory algorithm in quantum reinforcement learning for robotics,” IET Quantum Communication, 2024. [Online]. Available: https://doi.org/10.1049/qtc2.12111

[15] ——, “Quantum prioritized experience replay with madi-based priority and quantum circuit

mechanisms for optimizing reinforcement learning," International Journal of Advanced Technology and Engineering Exploration, vol. 11, no. 121, pp. 1664–1680, 2024. [Online].Available:http://dx.doi.org/10.19101/IJATEE. 2024.111100094

[16] K. Balasubramanian, T. Ganesan, K. Kalaivani, and S. Hemamalini, "Quantum-inspired hybrid machine learning framework for scalable computational next-generation technologies," International Journal of Computational Intelligence Systems, vol. 19, p. 92, 2026. [Online]. Available: https://doi.org/10.1007/s44196-026-01191-5

[17] D. Kong, J. Jin, X. Cui, J. Yan, J. Tang,Y. Fu, J. Xu, and D. Wang, "Cqnas: Convolutional quantum hybrid neural network architecture search for image classification," Neurocomputing, vol. 679, p. 133236, 2026. [Online]. Available: https://doi.org/10.1016/j.neucom.2026.133236

[18] L. M. Donaire, G. Ortega, F. Orts, G. E. M. Garzón, and E. Filatovas, "A hybrid quantum-classical approach for liver disease detection using quantum machine learning," Engineering Applications of Artificial Intelligence, vol. 164, p. 113240, 2026. [Online]. Available: https://doi.org/10.1016/j.engappai.2025.113240

[19] A. Verdone, F. Succetti, A. Ceschini, A. Rosato,A. Fioravanti, and M. Panella, "A hybrid quantum-neural network for heart disease classification," Biomedical Signal Processing and Control, vol. 113, p. 109185, 2026. [Online]. Available: https://doi.org/10.1016/j.bspc.2025.109185

[20] CERN Open Data Portal, "Cms outreach data: Wenu and wmunu samples," http://opendata.cern.ch, 2026, accessed 2026-04-19.