Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2026, 20(5), 473–486
https://doi.org/10.12913/22998624/217191
ISSN 2299-8624, License CC-BY 4.0

Received: 2025.12.15
Accepted: 2026.03.15
Published: 2026.04.01

Quantum-enhanced deep learning framework for real-time engagement recognition in e-learning environments

B. Vaidehi[1*], K. Arunesh[1]

[1] Department of Computer Science, Sri S. Ramasamy Naidu Memorial College (Affiliated to Madurai Kamaraj University), 626 203 Sattur, Tamil Nadu, India
* Corresponding author's e-mail: vaipri21@gmail.com

ABSTRACT

The rapid growth of online education has emphasized the importance of maintaining active student engagement to ensure effective learning outcomes. Traditional engagement detection methods use manual observation and they are time-consuming for large-scale datasets. This research presents a quantum-enhanced facial engagement recognition (Q-FER) system to overcome these constraints. This framework is inspired by the principles of quantum machine learning (QML). The suggested system merges the advantages of traditional deep learning (DL) and quantum computing (QC) by utilizing DL methods for feature extraction and employing a quantum convolutional neural network (QCNN) for classification. This approach utilizes quantum principles (QP) such as entanglement and superposition which enable the model to analyze numerous data patterns simultaneously. The dataset comprises 2.120 facial images of students captured during virtual classes. These images are classified into six emotional states, which are again categorized into two engagement levels: engaged and not engaged. The model employs Z-score normalization along with data augmentation methods to enhance generalization and minimize sensitivity to changes in lighting, orientation and facial expressions. Experimental results show that the Q-FER achieves a classification accuracy of 99.76%, significantly outperforming traditional convolutional neural networks (CNN). These results underscore the promise of quantum-assisted deep learning systems for analyzing real-time engagement in e-learning.

Keywords: quantum machine learning, facial engagement recognition, quantum convolutional neural network, deep learning, online education, student engagement.

INTRODUCTION

Nowadays, digital education has updated traditional practices and enabled students to learn beyond their geographical boundaries. This digital transition has introduced some challenges in assessing student engagement. It remains a key determinant of academic success and learning retention. Student engagement is commonly defined as the degree of attention, interest and active participation [1]. In virtual classrooms, teachers face the difficulty of evaluating engagement due to the lack of physical presence and non-verbal cues. Engagement monitoring is a crucial process because learners may demonstrate lower academic achievement and higher dropout rates [2]. Traditional approaches such as manual annotation are time-consuming and unsuitable for large-scale virtual classrooms. Scalable and automated engagement detection systems are increasingly essential [3] due to the expansion of online education. Machine learning (ML) has enabled the automation of engagement using behavioral and affective cues. Facial expression recognition (FER) is one of the most dependable techniques for evaluating emotional and involvement states [4]. FER-based systems analyze micro-expressions and facial dynamics to determine levels of concentration and emotional involvement. CNN and other DL models have achieved remarkable progress in emotion and engagement recognition but also face several critical limitations. Traditional CNNs necessitate extensive labeled datasets and substantial computational power to attain high accuracy. Their scalability is limited in real-time applications due to high training complexity, latency issues, and energy consumption [5]. Furthermore, the performance of CNNs may degrade under variations in illumination, camera angles, and partial occlusions common in real-world online learning settings [6]. These challenges have motivated the exploration of QC as a transformative paradigm for enhancing learning efficiency and computational performance. Quantum computing provides faster data processing compared to classical computation [7].

QML represents the intersection of QC and ML to enhance learning algorithms by exploiting quantum effects. QML algorithms such as QCNN have shown the ability to perform hierarchical feature extraction and classification more efficiently than traditional methods [8]. QCNN utilizes parameterized quantum circuits (PQC) to represent convolutional filters, allowing it to process high-dimensional data within quantum subspaces efficiently. Recent research has demonstrated that hybrid quantum -classical models outperform classical CNNs in tasks like image and emotion recognition, offering higher accuracy with reduced training parameters and faster convergence [9]. These advancements have inspired applications of QML in educational analytics and engagement prediction. Real-time engagement analysis requires models that are both computationally efficient and highly accurate. Although numerous DL architectures have been suggested for engagement detection, their reliance on traditional computation restricts scalability in real-time environments. On the other hand, QML, through its inherent parallelism, can accelerate training and inference, thereby reducing latency and resource utilization. Despite recent progress in QNN, there is a notable research gap in applying quantum convolutional neural networks to student engagement analysis. Current research predominantly examines traditional CNNs while insufficiently investigating how quantum models might improve the accuracy and efficiency of engagement detection. Addressing this gap forms the central motivation of this research. This proposed work introduces a Q-FER framework that integrates QCNN for efficient feature extraction and emotion -based engagement classification. The Q-FER framework processes facial image datasets representing various emotional states of students captured during online sessions. Each image is analyzed to determine engagement levels based on emotional cues. Experimental evaluations demonstrate that Q-FER surpasses classical CNN architectures in classification accuracy and computational cost. The major contributions of this work include:

1. A novel Q-FER architecture that integrates quantum – enhanced feature extraction for engagement detection.
2. The implementation of a QCNN model capable of using quantum entanglement to represent complex emotional patterns.
3. A comprehensive performance analysis comparing Q-FER with conventional CNN-based models on accuracy, training time, and scalability.

RELATED WORK

Monitoring student engagement has become a central focus in EDM especially in the context of Virtual Learning environments. Engagement is not only a measure of attention and participation but also an Emotional Indicator reflecting learners' motivation and cognitive involvement. Traditional methods such as self-reported surveys and manual observation are increasingly being replaced by AI-driven automated recognition systems. They analyze multimodal behavioral cues like facial expressions [10]. However, classical DL models such as CNN and RNN require high computational resources and large labeled datasets which constrain their scalability in real-time educational settings [11]. The advent of QML has created new opportunities for improving performance and computational efficiency. QCNN utilizes quantum superposition and entanglement to accelerate the processing of high-dimensional data compared to classical methods [7]. QML has shown remarkable potential in computer vision and image recognition tasks, suggesting that it could effectively optimize engagement detection frameworks when integrated with DL principles.

To provide a clear understanding of the existing literature, this section is divided into two areas. They are facial-based student engagement detection using classical DL models and quantum and hybrid ML models for image and emotion analysis. Tables 1 and 2 summarize key studies in each domain, outlining methodologies, datasets and major findings.

Facial-based student engagement detection

Facial expression analysis has been widely adopted as a primary modality for assessing student engagement in online and blended learning environments [12]. Researchers have developed CNN-based and multimodal models to identify Emotional cues linked to engagement, such as Happiness, Surprise, or Boredom. Table 1 presents selected studies that illustrate the evolution of DL techniques for engagement recognition.

Table 1 highlights major advancements in facial-based student engagement detection using deep learning techniques. Most studies employed CNN architectures to recognize emotional and behavioral cues from facial expressions, achieving accuracies between 85% and 95%. Alruwais and Zakariah [11] and Hsia et al. [13] confirmed the effectiveness of CNNs for binary engagement classification, while Hu and Gao [14] expanded analysis to multiple engagement dimensions. Integrating features like fatigue detection [15] and real-time video monitoring [16] improved contextual accuracy. However, existing models still face challenges such as high computational costs, small and imbalanced datasets, and sensitivity to lighting and camera conditions. These limitations underline the need for more efficient and scalable approaches, motivating the adoption of quantum-enhanced frameworks like the proposed Q-FER model.

Quantum and hybrid machine learning models

The field of QML has recently made substantial progress in the area of visual pattern recognition. By embedding Quantum Circuits into neural architectures, QML models utilize Quantum phenomena like superposition and entanglement [17] to enhance Feature Extraction. These models offer an advanced alternative to classical Deep Networks allowing more complex correlations to be captured using fewer parameters. Table 2 summarizes notable research studies that demonstrate the development and applications of QML models in image and emotion recognition tasks.

As summarized in Table 2, recent advancements in Quantum and Hybrid Neural Network architectures demonstrate significant improvements in efficiency, scalability, and emotional data processing. The fully parameterized QCNN [17] achieved 97.3% accuracy while reducing computational costs through Quantum Entanglement and optimized feature encoding. The Hybrid Quantum–Classical–Quantum CNN [9] further enhanced model speed and accuracy, validating the effectiveness of hybrid quantum-classical integration for image-based learning. Likewise, the Computation-Efficient QCNN [18] successfully optimized RGB-channel encoding for real-world visual inputs, minimizing qubit requirements and supporting large-scale deployment. In addition, the Parameterized Quantum Circuit [19] extended QML's application to Emotion Recognition, reducing model parameters by 50% while improving classification accuracy.

Collectively, these studies emphasize that Quantum-Enhanced Learning frameworks can effectively process complex, high-dimensional facial and emotional data. Such capabilities directly align with the needs of student engagement detection where real-time interpretation of emotional cues is critical. Therefore,

Table 1. Key studies on facial-based engagement detection

Ref.	Methodology	Dataset / Domain	Main findings & limitations
[11]	CNN trained on 40000 FER images categorized into engaged / disengaged states	Custom FER dataset	Effective for binary classification, though dataset imbalance limits generalization.
[13]	Deep CNN and regression model for emotional – behavioral engagement prediction	Online course video dataset	Showed strong correlation between facial features and Engagement and effective for static images but lacked temporal analysis and used a small dataset.
[14]	Deep CNN correlating facial expressions with six engagement metrics in synchronous online learning	Second-language (L2) online classes	Found strong correlation between facial emotions and affective engagement; restricted to emotional dimension.
[15]	FER combined with fatigue detection to assess learning status	E-learning video dataset	Identified fatigue as a negative engagement factor; requires high-resolution input for precision.
[16]	CNN for engagement detection during live online lectures	Real student lecture video streams	Proved effective in real-time online classes but was limited by lighting conditions, camera angles, and small dataset size.

Table 2. Recent research on quantum/hybrid neural models

Ref.	Model / Technique	Application / Dataset	Key outcomes
[18]	Fully parameterized QCNN with 2-qubit entangling gates	MNIST / Wine dataset	Attained 97.3% accuracy; quantum encoding reduced computational cost and training time.
[9]	Hybrid Quantum–Classical–Quantum CNN (QCQ-CNN)	Image classification benchmarks	Outperformed conventional CNNs in accuracy and speed, validating the hybrid model's efficiency.
[19]	Computation-efficient QCNN (CE-QCNN) for multi-channel inputs	KITTI autonomous driving dataset	Improved RGB-channel encoding efficiency and reduced qubit use for real-world visual tasks.
[20]	Parameterized Quantum Circuits (PQC) for emotion recognition	IEMOCAP / RECOLA / MSP-IMPROV datasets	Reduced parameters by 50% while improving emotion classification accuracy.

integrating Quantum feature extraction into engagement recognition through the proposed Quantum-Enhanced Facial Engagement Recognition framework has the potential to deliver more accurate and scalable solutions for modern E-learning environments.

METHODOLOGY

The proposed Q-FER framework is designed to efficiently identify and classify student engagement levels by integrating the strengths of DL and QML [21]. This architecture combines classical feature extraction using CNN with QCNN for final classification. This integration combines the advantages of Quantum Entanglement and Superposition which enable faster computation and enhanced pattern recognition capabilities. The Q-FER framework comprises seven major stages such as Input Acquisition, Preprocessing, Feature Extraction, Quantum Encoding, QCNN Classification, Engagement Prediction, and Evaluation. These phases are shown in Figure 1 and Algorithm 1 illustrates the complete process from input facial images to the creation of engagement predictions.

The dataset consists of 2,120 facial images of students captured during live online learning sessions. Each image corresponds to one of six observable emotional expressions. They are confused, engaged, frustrated, looking away, bored, and drowsy. These six emotion categories were mapped into two broader engagement levels such as engaged and not engaged. The engaged category includes confused, engaged and frustrated states which reflect cognitive involvement. Not

Input: Facial frames $F = \{F_1, F_2, \dots, F_n\}$
Output: $y^* \in$ {Engaged,Not Engaged}

1. Detect and crop facial ROI → resize to 128×128; convert to grayscale.
2. Normalize pixels: $Z = \frac{X-\mu}{\sigma}$.
3. Apply augmentation (rotation ±15°, flip, brightness, zoom).
4. Extract CNN features: $Y_{i,j} = (X * K)_{i,j}$; flatten → F_i.
5. Normalize $\| F_i \|^2 = 1$; encode as quantum state $| \psi_{in} \rangle = \sum_i \alpha_i \mid i \rangle$.
6. Apply gates $R_x(\theta_i), R_y(\theta_i)$, and CNOT for entanglement.
7. Transform: $| \psi_{out} \rangle = U(\theta) \mid \psi_{in} \rangle$.
8. Measure: $P_E = | \langle 0 \mid \psi_{out} \rangle |^2$, $P_N = | \langle 1 \mid \psi_{out} \rangle |^2$.
9. Predict: $y^* = \arg\max \{P_E, P_N\}$.
10. Optimize (W, θ)via Adam with loss
$$L = -\frac{1}{N} \sum [y \log(\hat{y}) + (1-y) \log(1-\hat{y})].$$

Algorithm 1. Quantum-enhanced facial engagement recognition (Q-FER) dataset description

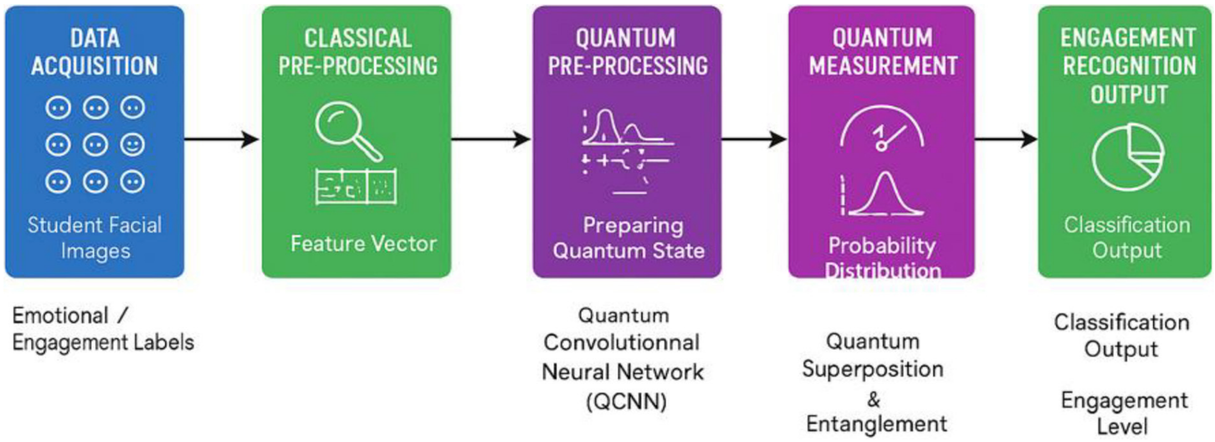

Figure 1. Flow diagram of the Quantum-Enhanced Facial Engagement Recognition (Q-FER) process

engaged category includes looking away, bored and drowsy states which signify behavioral disengagement. This mapping follows established educational psychology findings that distinguish between active engagement states and passive disengagement. A brief sensitivity analysis also confirmed that grouping confused and frustrated within the engaged class resulted in greater model stability and closer alignment with observed learning behaviors compared to treating them as independent categories.

Table 3 and Figure 2 present the distribution of the emotion categories and corresponding engagement levels. The dataset was divided into 80% for training and 20% for testing, maintaining an appropriate balance for DL based emotion recognition research. To enhance model generalization and robustness, several data augmentation techniques were applied which is including random rotation (±15°), horizontal flipping and controlled variations in brightness and zoom [22]. These augmentations help mitigate inconsistencies caused by lighting, facial orientation and camera positioning. Similar augmentation strategies have been used in prior educational emotion-recognition studies [21, 14] to improve model accuracy and resilience under diverse classroom conditions. The dataset structure and applied augmentations are illustrated in Figure 3.

Table 3. The number of images of each emotion in the dataset

Main class	Subclass	Count
Engaged	Confused	369
Engaged	Engaged (subclass)	347
Engaged	Frustrated	360
Not engaged	Looking away	423
Not engaged	Bored	358
Not engaged	Drowsy	263

Data preprocessing and feature extraction

The preprocessing stage serves as a fundamental step in ensuring that the facial images are standardized and suitable for subsequent feature extraction and quantum encoding. Initially, face detection is performed using the Haar Cascade Classifier [23], which identifies the Region of Interest (ROI) corresponding to each student's facial area. The detected faces are then cropped, resized to 128 × 128 pixels, and converted to grayscale to minimize computational complexity and reduce color-space redundancy [24]. Subsequently, data augmentation methods are utilized to improve the dataset's resilience and generalization ability. The augmentations include rotation (±15°), horizontal flipping, random brightness adjustment and zoom transformations which simulate variations in lighting, orientation, and camera position commonly encountered in real-world virtual classrooms [14]. These methods avoid overfitting and enhance the model's flexibility in various student settings.

To normalize pixel intensity values and standardize the data distribution, Z-score normalization is employed. This process ensures that all image features contribute equally to the learning process, effectively minimizing bias introduced by varying pixel intensities [25]. The normalization formula is given in Equation 1:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

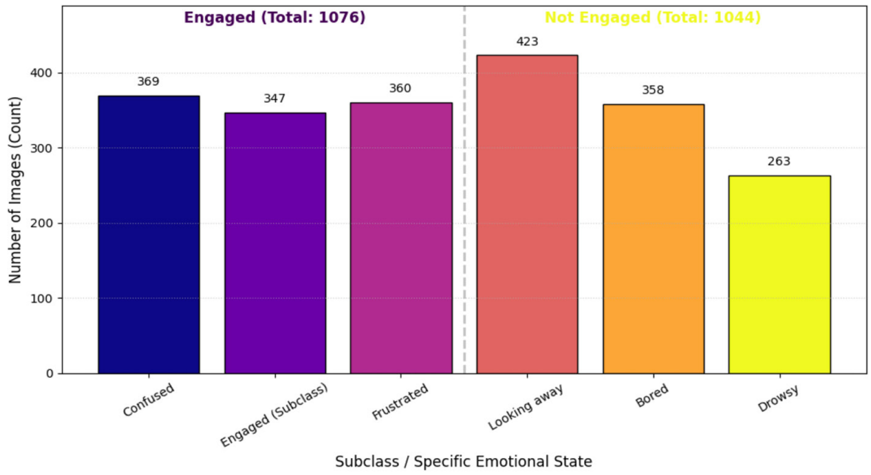

Figure 2. Dataset distribution

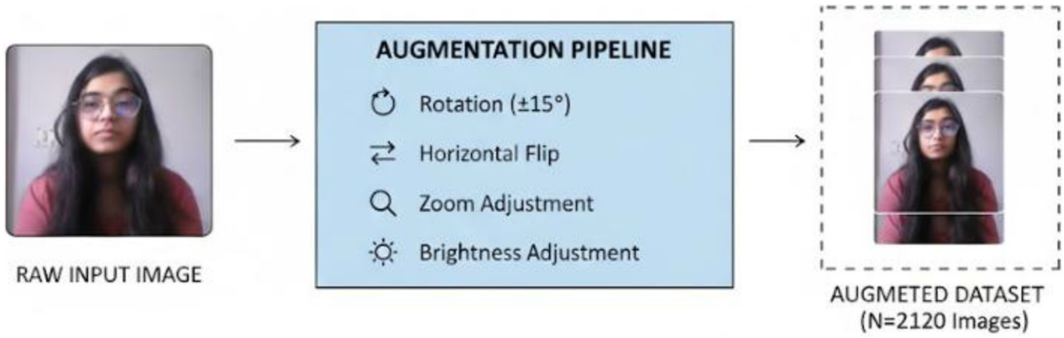

Figure 3. Data augmentation process

where: X represents the pixel intensity, μ is the Mean, and σ is the standard deviation.

CNN are used for feature extraction after normalization since they are renowned for extracting high-level spatial and semantic data. [26]. These features capture critical engagement cues such as eye gaze direction, head movement, and facial muscle dynamics, which form the emotional indicators of student engagement. The convolution operation can be mathematically represented in Equation 2:

$$Y_{i,j} = (X \times K)_{i,j} = \sum_{m} \sum_{n} X_{i+m,j+n} K_{m,n} \tag{2}$$

where: X is the input image matrix and K represents the convolutional kernel filter. The resulting feature maps are then flattened into numerical vectors that serve as input for the quantum encoding stage. Figure 4 illustrates the overall preprocessing and CNN-based feature extraction pipeline, showing the transformation from raw images through augmentation and feature learning to quantum encoding readiness.

Quantum encoding and QCNN architecture

The next stage involves transforming extracted features into quantum representations to control the computational advantages of quantum systems [21]. The numerical feature vectors obtained from the CNN are converted into quantum states through amplitude encoding which allows high-dimensional data to be efficiently represented within a limited number of qubits [18]. This transformation is mathematically expressed in Equation 3:

$$| \psi_{in} \rangle = \sum_{i=0}^{N-1} \alpha_i | i \rangle, \quad \text{where} \sum_{i} | \alpha_i |^2 = 1 \tag{3}$$

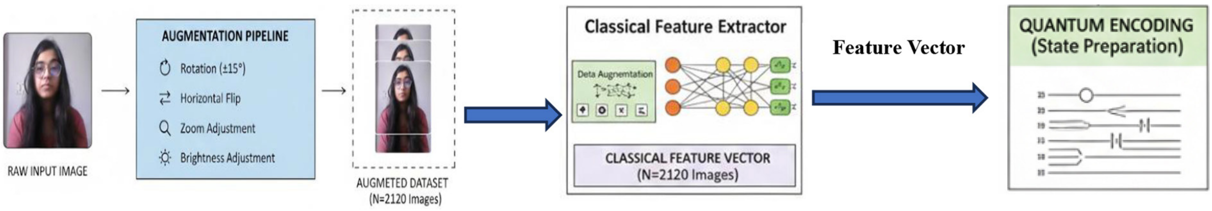

Figure 4. Preprocessing and CNN-based feature extraction pipeline

Here, α_i denotes the normalized amplitude corresponding to each feature value and $i\rangle$ represents the basis state of the qubit system [27]. This encoding enables parallel information processing through quantum superposition and allowing the model to explore multiple feature correlations simultaneously. The QCNN serves as the core of the proposed Q-FER framework. In this work, the QCNN is implemented with 8 qubits and a circuit depth of 4 layers where each layer comprises parameterized quantum gates Rx(θ) and Ry(θ) followed by CNOT gates to establish Entanglement among neighboring Qubits. Each rotation gate introduces a tunable parameter resulting in a total of 64 trainable quantum parameters. These parameters are optimized jointly with classical CNN weights during the training phase. The quantum transformation process is represented as Equation 4:

$$|\psi_{out}\rangle = U(\theta)\,|\psi_{in}\rangle \tag{4}$$

where: $U(\theta)$ is a unitary operator parameterized by the rotation angles θ.

After the transformation, measurements are performed. Then the corresponding expectation values are converted into classical probabilities. The measured output probabilities are interpreted as the likelihoods of the engaged and not engaged classes, respectively. Each QCNN layer performs a sequence of feature Entanglement, pooling and measurement operations. This allows efficient backpropagation through both Quantum and classical layers which combining representational depth with computational efficiency. Figure 5 illustrates the complete QCNN architecture and shows the qubit arrangement, Entanglement structure and measurement flow used for binary engagement classification.

Model training and evaluation

The Q-FER framework allows seamless integration of neural network components with QC simulations [30]. This approach combines the strengths of both classical and quantum computations [31]. The training procedure starts with the initialization of the quantum parameters of the QCNN denoted by rotation gate angles (θ). It is followed by classical weight adjustments via backpropagation employing the Adam optimizer with a learning rate (α = 0.001). The binary cross-entropy loss function is used to optimize the binary classification goal (engaged vs. not engaged). A 5-fold cross-validation approach was employed to ensure model generalization and minimize overfitting concerns. Each fold ensured that facial images from the same student did not appear in both training and testing sets, thereby preventing potential data leakage. Training was performed over 100 epochs with early stopping and dropout regularization applied to prevent overfitting once the validation accuracy plateaued. The reported performance represents the average accuracy across all folds and providing a more reliable estimate of the model's robustness. The loss function is expressed in Equation 5:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} y_i \log(\hat{y}_i) + (1-y_i) \\ \log(1-\hat{y}_i\,) \end{bmatrix} \tag{5}$$

where: y_i is the true class label and $\hat{y}_i$ is the predicted probability output from the QCNN measurement layer.

RESULTS AND DISCUSSION

The proposed Q-FER was implemented and tested using a dataset consisting of 2,120 facial images collected from students during online learning sessions. These images represented six fundamental emotional states which were categorized into two major engagement levels. In this work, 80% (1,697 images) were used for training and 20% (423 images) were used for testing and validation. During training, the Q-FER model

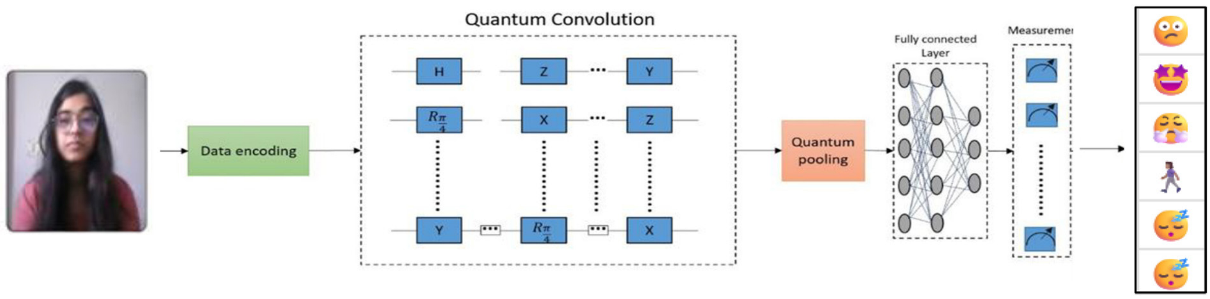

Figure 5. Quantum convolutional neural network architecture

demonstrated rapid and stable convergence. The training and validation accuracy and loss curves are shown in Figure 6. The plots reveal that both the training and validation losses decrease smoothly and indicating effective learning and minimal overfitting. Similarly, the accuracy curves for both sets increase steadily and converge near 99%, demonstrating consistent model behavior across epochs. This performance confirms that the model efficiently captured engagement-related facial features with the hybrid quantum layer enhancing nonlinear feature transformations and improving classification accuracy.

Quantitative evaluation

The effectiveness of the Q-FER framework was quantitatively evaluated using standard metrics, as shown in Table 4 and Figure 7. The exceptionally high accuracy indicates that the model correctly classified almost all engagement states. The perfect recall signifies that all engaged students were successfully identified, while the slightly lower precision suggests a very small number of false positive predictions. The overall F1-score confirms a balanced performance between precision and recall. The confusion matrix shown in Figure 8 supports these findings with most samples signifying correct classification for both engaged and not engaged categories. Only a few misclassifications occur, which can be attributed to overlapping emotional cues particularly between "confused" and "frustrated" or "bored" and "drowsy" expressions. This slight misunderstanding is anticipated in practical emotion recognition tasks and does not notably impact the model's dependability.

Visual prediction analysis

The effectiveness of the proposed framework is illustrated through Figure 9 and presents sample predictions from the validation set. Each image is labeled with its actual and predicted engagement level. The Q-FER model demonstrates excellent prediction accuracy across various lighting conditions, facial orientations and expression

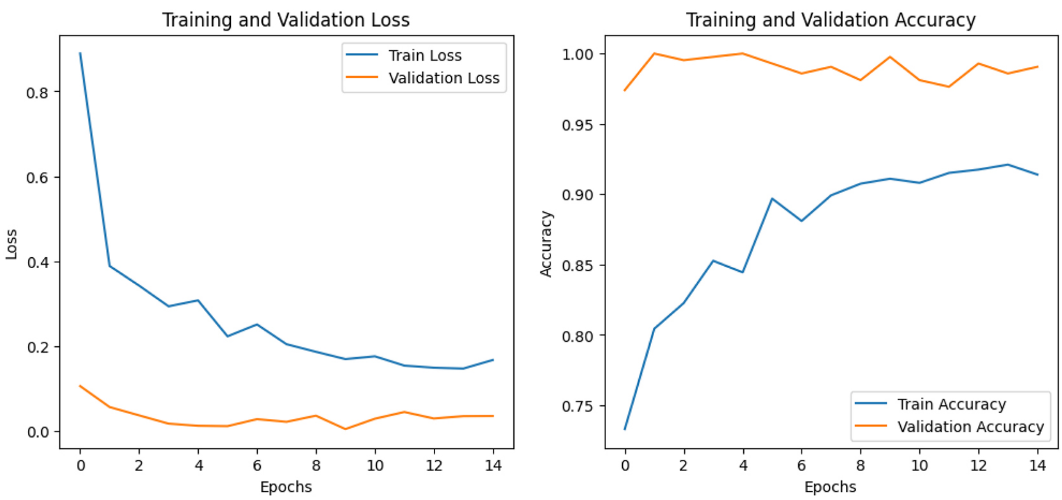

Figure 6. Training and validation accuracy and loss curves of the Q-FER framework

Table 4. Performance metrics

Metric	Score
Accuracy	99.76%
Precision	99.52%
Recall	100.00%
F1-Score	99.76%

intensities. Students exhibiting focused gaze, attentive posture or expressive facial cues were classified as Engaged whereas those showing a downward gaze or fatigue were classified as not engaged. This visual analysis highlights the model's ability to recognize facial differences for real-time monitoring applications in virtual classrooms.

Comparative analysis with existing models

A comparative analysis was carried out to evaluate the performance of the proposed Q-FER framework against a baseline classical CNN and DL architectures. The baseline CNN used for comparison had four convolutional layers with 3 × 3 kernels and ReLU activation. Each convolutional layer was followed by a max-pooling and a batch-normalization layer. The network included two fully connected layers with 128 and 64 neurons and a final softmax output layer for classification. The model was trained for 100 epochs using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss. To maintain consistency, all models adopted the same preprocessing pipeline and image resolution of 128 × 128 pixels. The results are summarized in Table 5 and illustrated in Figure 10. The results show that the baseline CNN achieved a satisfactory accuracy of 92.15% and the Q-FER model considerably improved recognition accuracy by effectively learning nonlinear relationships among facial features related to engagement levels.

In addition to the baseline model, modern architectures such as ResNet-50, MobileNet V2, and EfficientNet-B0 were implemented. Although these networks demonstrated strong performance, the proposed Q-FER framework attained the highest overall accuracy. A further comparison with existing quantum CNN frameworks is presented in Table 6 and indicates that Q-FER also surpasses earlier quantum approaches in all metrics. As shown in Figure 11, the Q-FER model demonstrates superior generalization and robustness in engagement recognition. The comparative results confirm that the proposed Q-FER framework consistently outperforms both conventional and modern DL models as well as previously reported quantum CNNs.

Runtime benchmark analysis

The runtime performance of the proposed Q-FER framework was compared with classical and modern CNN architectures to evaluate computational efficiency. Three major metrics were analyzed are inference latency per frame,

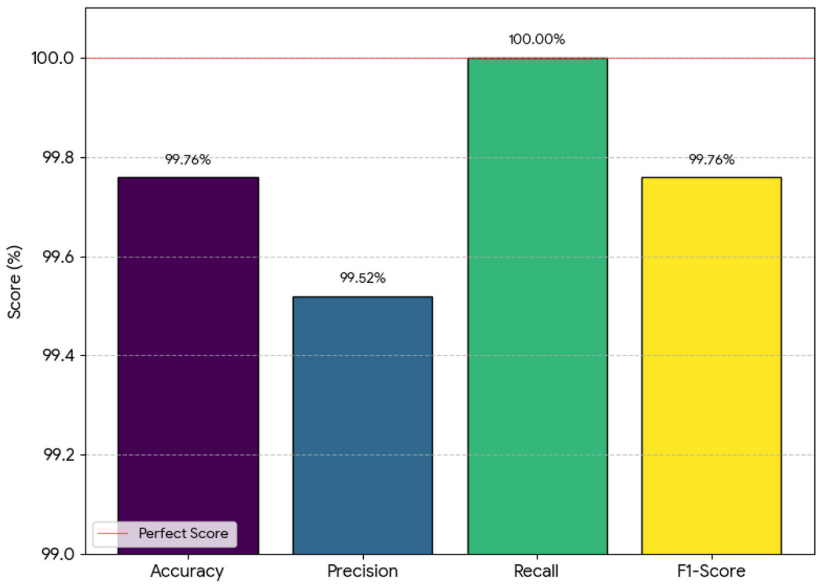

Figure 7. Performance metrics of the Q-FER framework

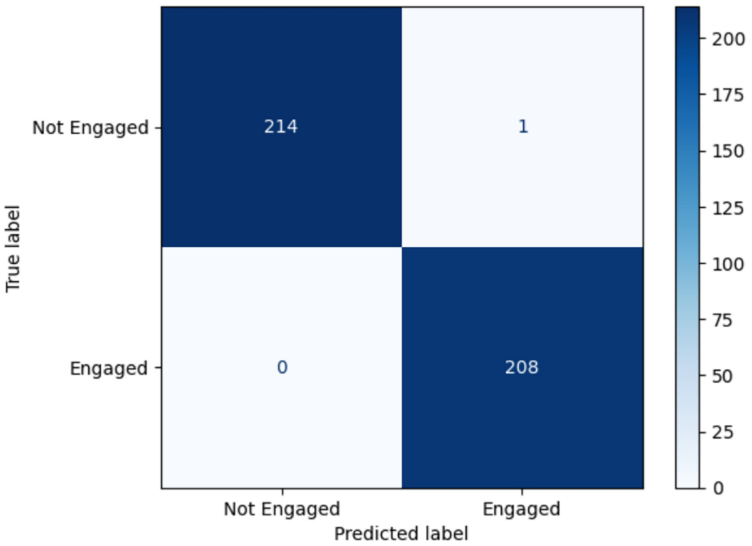

Figure 8. Confusion matrix of the Q-FER modl

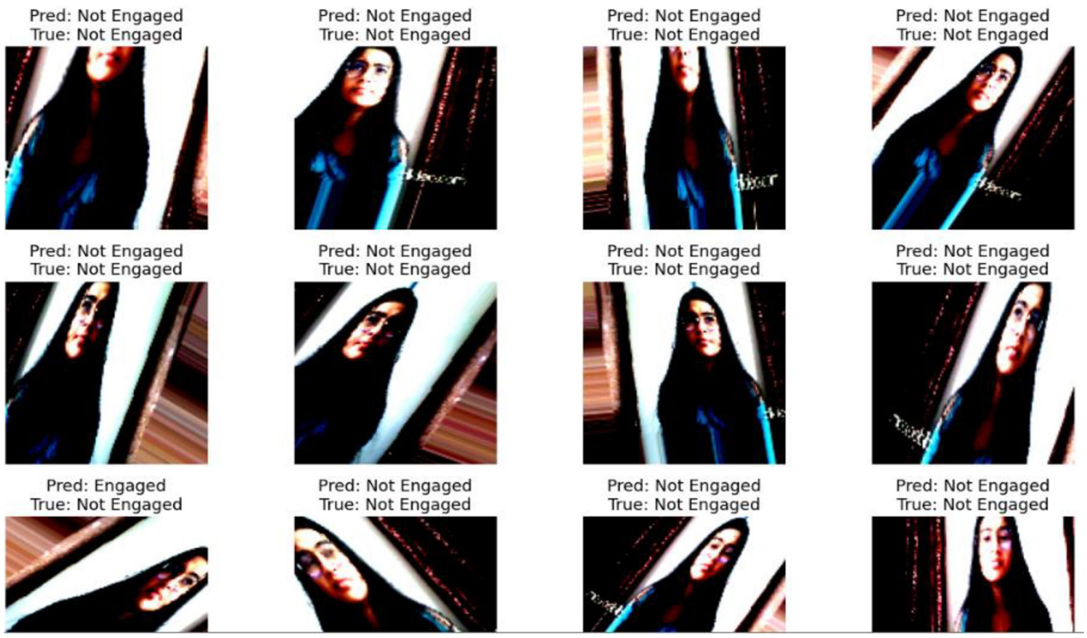

Figure 9. Sample visual predictions with true and predicted labels

throughput (frames per second) and training time (in hours). The results are presented in Table 7.

As shown in Table 7 and Figure 12, the proposed Q-FER model demonstrates slightly higher inference latency (22 ms/frame) compared with conventional CNNs due to the quantum circuit simulation process. However, it maintains real-time throughput of approximately 45 fps, which is adequate for continuous engagement monitoring. The training duration (1.8 hours) is lower than that of deeper models such as ResNet-50 (2.4 h) and EfficientNet-B0 (2.1 h), confirming its computational efficiency. The bar chart in Figure 12 further highlights Q-FER's favorable balance between latency, throughput, and training

Table 5. Comparative analysis of CNN vs Q-FER

Model	Accuracy (%)
Classical CNN	92.15
ResNet-50	96.83
MobileNet V2	95.47
EfficientNet-B0	97.12
Proposed Q-FER	99.76

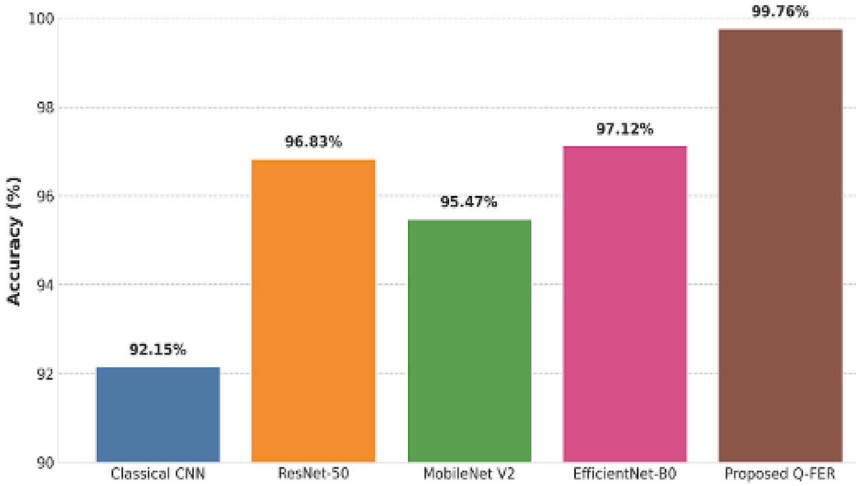

Figure 10. Comparative analysis of CNN vs Q-FER

Table 6. Comparative summary of quantum models

Ref.	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[17]	Fully parameterized QCNN	97.3	96.9	97.1	97.0
[9]	Hybrid QCQ-CNN	98.5	98.1	98.3	98.2
[18]	CE-QCNN	96.1	95.4	96.0	95.6
[19]	PQC-based emotion recognition	80.85	81.2	80.1	80.6
	Proposed (Q-FER)	99.76	99.52	100.00	99.76

Table 7. Runtime benchmark comparison between baseline and proposed models

Model	Inference latency (ms/frame)	Throughput (fps)	Training time (hours)
Classical CNN	12.3	81.3	1.0
ResNet-50	18.6	53.7	2.4
MobileNet V2	14.8	67.5	1.6
EfficientNet-B0	16.9	59.2	2.1
Proposed Q-FER	22.0	45.4	1.8

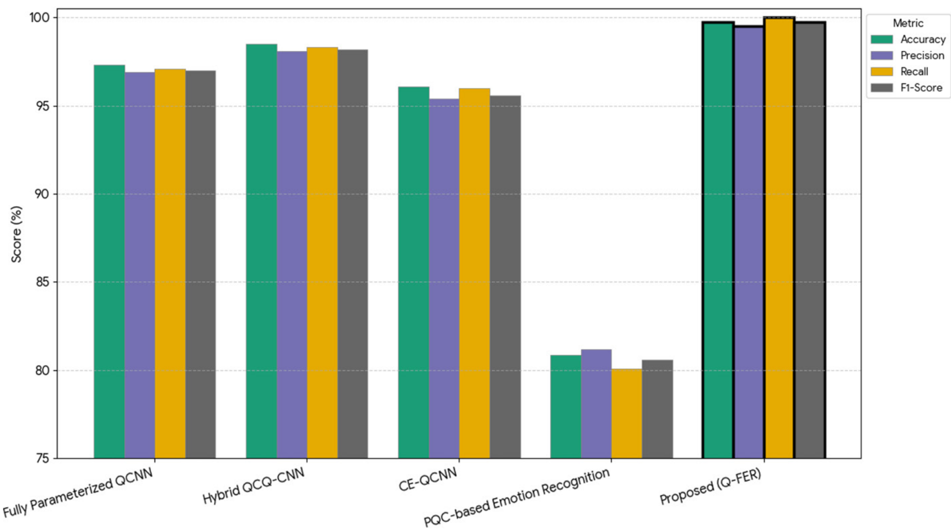

Figure 11. Comparison of Q-FER vs. existing quantum-CNN frameworks

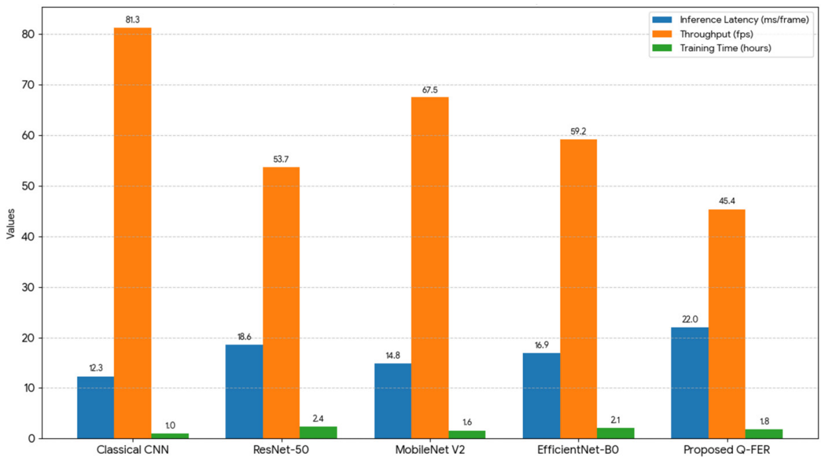

Figure 12. Runtime benchmark comparison of baseline vs. proposed models

time. Although minor computational overhead is observed, Q-FER achieves the highest classification accuracy (99.76%) with reduced parameter complexity, demonstrating that the hybrid quantum–classical design provides an effective trade-off between performance and runtime cost for real-time e-learning applications.

CONCLUSIONS

The proposed Q-FER framework integrates the capabilities of DL and QML to classify student engagement levels in online learning environments. By combining classical feature extraction with a hybrid quantum inspired classification network, the model effectively captures complex emotional and behavioral cues associated with engagement. Using a dataset of 2.120 facial images, the Q-FER model achieved an accuracy of 99.76% which is demonstrating strong performance and robustness. The Q-FER framework achieved superior classification accuracy and more efficient feature representation compared with conventional CNN architectures. However, since the present implementation is based on quantum circuit simulations, the results reflect representational and learning advantages rather than experimentally verified computational speedups or scalability. Future studies will implement Q-FER on real Quantum hardware to empirically evaluate runtime performance and computational efficiency. Overall, Q-FER illustrates the potential of Quantum enhanced learning for emotion and engagement recognition in digital education. By enabling automated identification of disengaged learners, it provides a foundation for adaptive and personalized learning environments. Future research will also extend this framework to include multimodal features such as eye gaze, voice tone, and body posture, and to integrate it with Learning Management Systems to promote active and student-centered online learning.

REFERENCES

1. Fredricks J. A., Blumenfeld P. C., and Paris A. H., School engagement: Potential of the Concept, State of the Evidence, Rev Educ Res, Mar. 2004; 74(1): 59–109. https://doi.org/10.3102/00346543074001059
2. Wakjira A. and Bhattacharya S., Predicting student engagement in the online learning environment, International Journal of Web-Based Learning and Teaching Technologies, Oct. 2021; 16(6): 1–21, https://doi.org/10.4018/IJWLTT.287095
3. Hu M. and Li H., Student engagement in online learning: A review, in 2017 International Symposium on Educational Technology (ISET), IEEE, Jun. 2017; 39–43. https://doi.org/10.1109/ISET.2017.17
4. Calvo R., D'Mello S., Gratch J., Kappas A., D'Mello S., and Graesser A. C., Feeling, Thinking, and computing with affect-aware learning technologies, in The Oxford Handbook of Affective Computing, Oxford University Press, 2015. https://doi.org/10.1093/oxfordhb/9780199942237.013.032
5. Thuong D. T. M., Huy N. P., Phung T.-N., and Cuong D. N., A Lightweight deep learning network for emotion recognition applications on portable devices, International Journal of Knowledge and Systems Science, Apr. 2025; 16(1): 1–23, https://doi.org/10.4018/IJKSS.373712
6. Gupta S., Kumar P., and Tekchandani R. K., Facial emotion recognition based real-time learner

engagement detection system in online learning context using deep learning models, Multimed Tools Appl, Mar. 2023; 82(8): 11365–11394, https://doi.org/10.1007/s11042-022-13558-9

7. Senokosov A., Sedykh A., Sagingalieva A., Kyriacou B., and Melnikov A., Quantum machine learning for image classification, Mach Learn Sci Technol, Mar. 2024; 5(1): 015040, https://doi.org/10.1088/2632-2153/ad2aef
8. Napoli F., Campanile L., De Gregorio G., and Marrone S., Quantum Convolutional Neural Networks for Image Classification: Perspectives and Challenges, 2025, https://doi.org/10.5220/0013521500003944
9. Long C., Huang M., Ye X., Futamura Y., and Sakurai T., Hybrid quantum-classical-quantum convolutional neural networks, Scientific Reports, Aug. 2025, 15(1): 31780, https://doi.org/10.1038/s41598-025-13417-1
10. Gupta S., Kumar P., and Tekchandani R., A multimodal facial cues based engagement detection system in e-learning context using deep learning approach, Multimed Tools Appl, Jul. 2023; 82(18): 28589–28615, https://doi.org/10.1007/s11042-023-14392-3
11. Alruwais N. and Zakariah M., Detecting Student Engagement with Convolution Neural Network and Facial Expression Recognition, Traitement du Signal, Apr. 2025; 42(2), https://doi.org/10.18280/ts.420229
12. Qarbal I., Sael N., and Ouahabi S., Enhancing emotion recognition model for a student engagement use case through transfer learning, IAES International Journal of Artificial Intelligence (IJ-AI), Apr. 2025; 14(2): 1576, https://doi.org/10.11591/ijai.v14.i2.pp1576-1586
13. Hsia C.-H., Chiang B., Ke L.-Y., Ciou Z.-Y., and Lai C.-F., Student Engagement Analysis Using Facial Expression in Online Course, in 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA), IEEE, Oct. 2022; 1–2. https://doi.org/10.1109/IET-ICETA56553.2022.9971565
14. Hu X. and Gao J., Facial expression recognition reveals students' engagement in online class: Correlations with six engagement measurements, PLoS One, Oct. 2025; 20(10): e0334232, https://doi.org/10.1371/journal.pone.0334232
15. Ding X. and Mariano V. Y., Learning status recognition method based on facial expressions in e-learning, Journal of Advanced Computational Intelligence and Intelligent Informatics, Jul. 2024; 28(4): 793–804, https://doi.org/10.20965/jaciii.2024.p0793
16. Siswantoro J., Rahmadiarto J., and Naufal M. F., Facial expression recognition to detect student engagement in online lectures, Teknika, Jun. 2024; 13(2): 226–232, https://doi.org/10.34148/teknika.v13i2.853
17. Quantum Computing: The Future of Information Processing - 1st Edition. Accessed: Dec. 10, 2025. [Online]. Available: https://www.routledge.com/Quantum-Computing-The-Future-of-Information-Processing/Tiwari-KumarTyagi-Nagaraj/p/book/9781032813684
18. Prajapat S., Tomar M., Kumar P., Kumar R., and Vasilakos A. V., Quantum computing meets deep learning: A QCNN model for accurate and efficient image classification, Mathematics, Oct. 2025; 13(19): 3148, https://doi.org/10.3390/math13193148
19. Roh E. J., Im C., Jeong W., and Park S., Computation-efficient quantum convolutional neural networks for autonomous driving applications, The Journal of Supercomputing 2025; 81(8): 1033, https://doi.org/10.1007/S11227-025-07507-0
20. [20] Rajapakshe T., Rana R., Riaz F., Khalifa S., and Schuller B. W., Representation Learning with Parameterised Quantum Circuits for Advancing Speech Emotion Recognition, Jan. 2025, Accessed: Nov. 18, 2025. https://arxiv.org/pdf/2501.12050
21. Quantum Machine Learning: A Modern Approach - 1st Edition - S Karthike. Accessed: Dec. 10, 2025. [Online]. Available: https://www.routledge.com/Quantum-Machine-Learning-A-Modern-Approach/Karthikeyan-Akila-Sumathi-Poongodi/p/book/9781032544717
22. Mahendra H. N., Pushpalatha V., Mallikarjunaswamy S., Rama Subramoniam S., and Praveen J., Quantum convolutional neural network-based hybrid network for remote sensing image classification, Eng Appl Artif Intell, Jan. 2026; 164, https://doi.org/10.1016/J.ENGAPPAI.2025.113195
23. Jayaswal R., Mohd. Ansari A., Dixit M., Singh D. K., and Ahmad S., Advances in facial expression recognition technologies for emotion analysis, Discover Computing, Sep. 2025; 28(1): 203, https://doi.org/10.1007/s10791-025-09699-8
24. Liu J., Investigating Data Augmentation Strategies for Computer Vision Facial Expression Recognition, May 01, 2023. https://doi.org/10.58445/rars.192
25. Recent Advances in Computational Methods in Science and Technology: 2-. Accessed: Dec. 10, 2025. https://www.routledge.com/Recent-Advances-in-Computational-Methods-in-Science-and-Technology-2-Volume-Set/Kaur-Kaur-Kumar/p/book/9781041243595
26. Sandler M., Howard A., Zhu M., Zhmoginov A., and Chen L.-C., MobileNetV2: Inverted Residuals and Linear Bottlenecks, in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Jun. 2018; 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

27. Li Y., Qi Z., Li Y., Yang H., Shang R., and Jiao L., Quantum splitting convolutional neural network-based distributed quantum disease detection model, Neurocomputing, Nov. 2025; 655: 131410, https://doi.org/10.1016/j.neucom.2025.131410
28. Fan F., Shi Y., Guggemos T., and Zhu X. X., Hybrid Quantum-Classical Convolutional Neural Network Model for Image Classification, IEEE Trans Neural Netw Learn Syst, Dec. 2024; 35(12): 18145–18159, https://doi.org/10.1109/TNNLS.2023.3312170
29. Ahmed T., Kashif M., Marchisio A., and Shafique M., A comparative analysis and noise robustness evaluation in quantum neural networks, Sci Rep, Sep. 2025; 15(1): 33654, https://doi.org/10.1038/s41598-025-17769-6
30. Broughton M. et al., TensorFlow Quantum: A Software Framework for Quantum Machine Learning, Andrea Skolik, Mar. 2020; 12: 23, Accessed: Nov. 21, 2025. https://arxiv.org/pdf/2003.02989
31. Khang A. and Ratha K., The Quantum Evolution : Application of AI and Robotics in the Future of Quantum Technology, 2025; 520.
32. Heaton J., Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning, Genet Program Evolvable Mach, Jun. 2018; 19(1–2): 305–307, https://doi.org/10.1007/s10710-017-9314-z