International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 1502-1518
https://doi.org/10.70917/ijcisim-2026-4484
Article

Performance Evaluation of Quantum Machine Learning Models for Cyberbullying Detection

Vandana P. Pawar [1*], Prashant M. Yawalkar [2]

[*1]Department of Computer Engineering, MET's Bhujbal Knowledge city, Institute of Engineering, Nashik, Maharashtra, India, vandanapawarp@gmail.com ORCID ID: 0009-0004-4913-9668
[2]Department of Computer Engineering, MET's Bhujbal Knowledge city, Institute of Engineering, Nashik, Maharashtra, India, prashant25yawalkar@gmail.com ORCID ID: 0000-0002-4174-5813

Abstract: Cyberbullying on social media has become a serious issue that affects users' mental well-being, safety, and online engagement. Detecting cyberbullying automatically is challenging due to informal writing styles, short messages, and the presence of subtle abusive patterns that often overlap with normal communication. This paper presents a comparative evaluation of quantum-based learning models for cyberbullying detection using a unified experimental pipeline. The proposed system applies text preprocessing followed by TF-IDF feature extraction and PCA-based dimensionality reduction, and then performs classification using Quantum Support Vector Machine (QSVM), Quantum Neural Network (QNN), Quantum Natural Language Processing (QNLP), and Quantum Singular Value Decomposition (QSVD). In addition, a hybrid QNN+QSVM ensemble is evaluated to examine whether combining complementary decision mechanisms improves prediction consistency. Experimental results show that QSVM achieved the best overall accuracy of 88.15%, with 92.52% precision, 92.23% recall, and an F1-score of 92.38%. QSVD delivered comparable performance with 88.08% accuracy, 93.53% precision, 90.98% recall, and an F1-score of 92.24%, indicating strong class separation with reduced feature complexity. QNN and QNLP produced competitive results, achieving 84.12% and 83.85% accuracy, respectively, with F1-scores close to 90%, demonstrating stable detection capability under quantum embedding. The hybrid QNN+QSVM model achieved 87.37% accuracy and an F1-score of 91.87%, showing balanced precision (92.06%) and recall (91.68%) and confirming that ensemble integration can improve reliability over individual quantum neural approaches. These findings suggest that kernel-driven quantum classifiers and decomposition-based quantum representations can provide strong performance for cyberbullying detection while maintaining consistent predictive behavior across evaluation measures.

Keywords: Cyberbullying Detection, Quantum Machine Learning, QSVM, QNN, QNLP, SVD, Hybrid Ensemble (QNN+QSVM), Social Media Text Classification

1. Introduction

Social media platforms have become a primary medium for communication, content sharing, and public discussion. Along with their benefits, these platforms have also enabled the rapid spread of harmful behavior, including cyberbullying. Cyberbullying refers to repeated or intentional online actions that humiliate, harass, threaten, or emotionally harm an individual or a group. Unlike traditional bullying, online harassment can occur anonymously, reach a wider audience instantly, and persist through reposts and digital permanence. As a result, cyberbullying has been linked with psychological distress, reduced academic performance, and long-term emotional consequences. Manual monitoring and reporting mechanisms are insufficient for controlling cyberbullying at scale. The volume of user-generated content is extremely high, and abusive messages are often written using informal language, abbreviations, spelling variations, sarcasm, or coded terms. In many cases, cyberbullying is also context-dependent: a sentence may appear normal in isolation but becomes harmful when combined with conversational history, targeted identity, or repeated patterns of interaction. These factors make automatic detection both necessary and challenging.

Over the last decade, cyberbullying detection has been studied using machine learning and deep learning techniques. Traditional classifiers such as Support Vector Machines, Naive Bayes, and Random Forests have been widely adopted because they are efficient and work well with engineered textual features. More recently, deep learning models such as CNNs, RNNs, and transformer-based architectures have improved detection performance by learning richer semantic patterns from text. However, these models still face limitations, including large computational requirements, sensitivity to dataset imbalance, and difficulty in generalizing when language styles change across platforms or regions.

Quantum machine learning (QML) has emerged as an alternative computational direction for classification problems. In QML, mathematical operations such as kernel construction, feature mapping, and optimization can be modelled using quantum principles. Although practical quantum computing is still evolving, quantum-inspired methods and hybrid quantum-classical systems are now being explored for real-world learning tasks. In text classification, QML may offer benefits through different feature transformations and decision boundaries compared with conventional models, particularly in settings where patterns are complex and non-linear.

The study investigates the suitability of quantum-based learning models for detecting cyberbullying in social media text. A consistent experimental pipeline is used, including TF-IDF representation and PCA-based dimensionality reduction, followed by model training and evaluation. The study considers four quantum models: Quantum Support Vector Machine (QSVM), Quantum Neural Network (QNN), Quantum Natural Language Processing (QNLP), and Quantum Singular Value Decomposition (QSVD), and also evaluates a hybrid ensemble combining QNN and QSVM. The performance is compared using accuracy, precision, recall, and F1-score. The reported results indicate that QSVM provides the strongest standalone performance, while the hybrid design improves consistency over multiple models, supporting the potential of QML frameworks for cyberbullying detection.

The remaining sections of the paper are organized as follows: Section 2 discusses related studies in cyberbullying detection and quantum learning methods. Section 3 describes the proposed methodology, including preprocessing, feature extraction, and model construction. Section 4 presents the experimental setup and the obtained results. Section 5 provides a discussion of findings and comparisons. Finally, Section 6 concludes the study and outlines future research directions.

2. Literature Review

2.1 Classical Approaches

Classical machine learning techniques have formed the foundation of cyberbullying detection research, with several studies investigating both conventional classifiers and deep learning models for identifying abusive content on social media platforms. Support Vector Machine (SVM) and Naive Bayes (NB) have been extensively applied to sentiment analysis-based cyberbullying detection on Twitter (currently X) [11]. By utilizing n-gram features, particularly the 4-gram representation, the SVM classifier achieved an accuracy of 92.02%, whereas NB attained 81.1%, indicating that SVM provides better discrimination of textual patterns when appropriate feature representations are employed.

To improve interpretability, an explainable multitask framework named mExCB was introduced for detecting cyberbullying in code-mixed social media content [12]. The framework combines Convolutional Neural Networks (CNNs) with Gated Recurrent Units (GRUs) and incorporates attention mechanisms at both the word and sub-sentence levels. In addition to cyberbullying detection, the model performs rationale generation, target identification, and sentiment analysis within a unified architecture. Experimental evaluation reported improvements of 2.88% in detection accuracy and 4.24% in rationale identification. Nevertheless, the framework exhibited reduced effectiveness when processing implicit or sarcastic expressions.

Recurrent Neural Network (RNN)-based models have also been investigated for sentiment-aware cyberbullying detection [13]. These approaches capture the sequential dependencies within textual conversations and determine message polarity to identify harmful interactions. Their ability to process contextual information makes them suitable for monitoring conversations in messaging applications such as WhatsApp.

A comparative analysis involving XGBoost, K-Nearest Neighbors (KNN), Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), and Deep Neural Networks (DNN) demonstrated the benefits of incorporating sentiment and emotion-related features into the classification process [14]. The integration of emotional and contextual information significantly enhanced predictive performance, with the best-performing model achieving an accuracy of 98.90%. These findings emphasize the importance of combining informative feature engineering with robust classification algorithms.

Data augmentation has also been adopted to improve the generalization capability of cyberbullying detection systems. A framework integrating Easy Data Augmentation (EDA), sentiment analysis, and SVM for filtering abusive Instagram comments achieved an accuracy of 92.5% while increasing cross-validation performance by 2.52% [15]. The results suggest that augmentation strategies can effectively improve model robustness, particularly when training data are limited.

Deep learning approaches have further demonstrated superior performance in comparison with conventional machine learning techniques. A comparative study employing SVM, NB, and CNN models with n-gram features for Twitter-based cyberbullying detection reported that CNN achieved the highest accuracy of 94.43%, outperforming SVM (91.64%) and NB (81.1%) [16]. This improvement highlights the capability of deep neural networks to automatically extract complex semantic and contextual features from textual data.

2.2 Quantum Approaches

Quantum machine learning has recently emerged as a promising research direction for addressing complex optimization and classification problems by exploiting quantum computational principles. Variational Quantum Eigensolvers (VQEs) utilizes parameterized quantum circuits to approximate eigenvalues and efficiently solve optimization problems, providing a practical foundation for quantum-enhanced learning algorithms [17], [18].H1: Servant leadership has a positive effect on the Human Capital Readiness of Kodim units in the national food-security program.

Among quantum learning paradigms, Quantum Neural Networks (QNNs) have attracted significant research interest because of their ability to integrate quantum phenomena into neural computation. Narayanan and Menneer [19] presented one of the earliest theoretical formulations of QNNs, demonstrating that quantum circuits could emulate the functionality of classical neural networks. Building upon this concept, Cao et al. [20] introduced quantum neurons with threshold activation mechanisms that exploit quantum properties such as entanglement and coherence during the learning process. Subsequently, Schuld et al. [21] proposed quantum perceptron models capable of simulating activation functions, thereby extending the applicability of quantum neural architectures to supervised classification tasks.

Several studies have focused on improving the efficiency and scalability of QNNs. Ricks and Ventura [22] developed training strategies based on minimally entangled qubits to reduce computational complexity while maintaining classification capability. To enhance learning in high-dimensional feature spaces, Panella and Martinelli [23] investigated nonlinear quantum circuits for feature mapping and proposed Quantum Neuro-Fuzzy Networks that combine quantum computation with fuzzy reasoning to achieve more effective decision-making.

Research has also explored quantum memory mechanisms to enhance learning and information retrieval in quantum systems. Ventura and Martinez [27] introduced a quantum associative memory model that demonstrated higher storage capacity than conventional associative memories. Trugenberger further improved these models by incorporating probabilistic retrieval mechanisms, resulting in enhanced memory recall performance [28], [29]. In another significant contribution, Silva et al. [33 - 35] proposed the Superposition-based Learning Algorithm (SLA), which integrates Quantum Random Access Memory (QRAM) with weightless neural networks to enable simultaneous learning and decision-making through quantum superposition principles.

3. Methodology

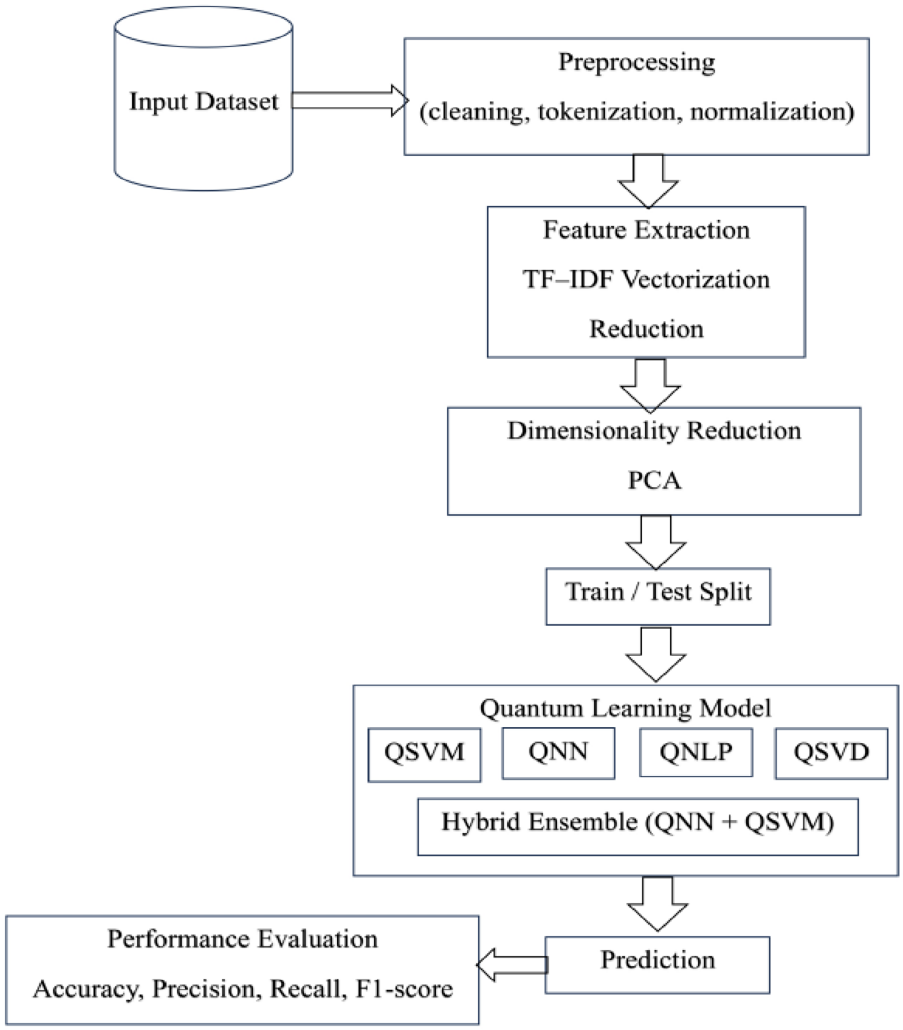
3.1 Proposed System

This section presents the proposed quantum machine learning-based framework developed for cyberbullying detection from social media text. The proposed system follows a unified processing pipeline so that all models are trained and evaluated under identical settings. The workflow consists of five major stages:

(i) Text preprocessing,
(ii) TF–IDF feature extraction,
(iii) Dimensionality reduction using PCA,
(iv) Quantum model learning (QSVM, QNN, QNLP, QSVD), and
(v) Hybrid decision making using an ensemble of QNN and QSVM.

The overall objective is to learn a mapping from a text sample to its corresponding class label (cyberbullying or non-cyberbullying) using quantum-inspired learning mechanisms.

Fig. 1: System Architecture



3.2 Problem Definition

Let the labelled dataset be expressed as:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{1}$$

where x_i represents the i^{th} text sample and $y_i \in \{0,1\}$ denotes the corresponding class label.

In this study, $y_i = 1$ represents cyberbullying content,

whereas $y_i = 0$ represents normal content. The objective is to learn a classifier $f(\cdot)$ such that:

$$\hat{y}_i = f(x_i), \tag{2}$$

where $\hat{y}_i$ denotes the predicted class label for the i^{th} text sample. The classifier is trained such that $\hat{y}_i$ matches the true label y_i with maximum correctness under standard evaluation metrics.

3.3 Text Preprocessing

Raw social media posts often contain noisy elements such as hyperlinks, special characters, repeated symbols, and inconsistent casing. To ensure stable feature representation, each text sample x_i is transformed using a preprocessing function $\phi(\cdot)$:

$$x_i' = \phi(x_i), \tag{3}$$

Where x_i' denotes the cleaned text. The preprocessing stage aims to standardize the input while preserving key linguistic cues that are useful for distinguishing harmful language from normal communication.

3.4 TF–IDF Feature Extraction

After preprocessing, each text sample is converted into a numerical representation using Term Frequency–Inverse Document Frequency (TF–IDF). Let t be a term in the vocabulary $\mathcal{V}$ and let d denote a document. The term frequency is computed as:

$$\mathrm{tf}(t,d) = \frac{n(t,d)}{\sum_{t' \in \mathcal{V}} n\,(t',d)} \tag{4}$$

where $n(t,d)$ represents the number of occurrences of term t in document d. The inverse document frequency is defined as:

$$\mathrm{idf}(t) = \log\left(\frac{N}{1+\mathrm{df}(t)}\right) \tag{5}$$

where $\mathrm{df}(t)$ denotes the number of documents containing term t, and N is the total number of documents. The TF–IDF weight is computed as:

$$w(t,d) = \mathrm{tf}(t,d) \cdot \mathrm{idf}(t) \tag{6}$$

Thus, each preprocessed sample x_i' is mapped into a vector:

$$\mathbf{v}_i = [w(t_1, x_i'), w(t_2, x_i'), \ldots, w(t_m, x_i')] \in \mathbb{R}^m \tag{7}$$

where $m = |\mathcal{V}|$ is the vocabulary size. TF–IDF provides a sparse representation that highlights words that are frequent in a particular post but uncommon across the dataset, making it suitable for cyberbullying text classification.

3.5 Dimensionality Reduction using PCA

TF–IDF vectors are typically high-dimensional, which increases training time and can introduce redundancy. To obtain compact representations, Principal Component Analysis (PCA) is applied.

Let $\mathbf{V} \in \mathbb{R}^{N\times m}$ be the TF–IDF matrix formed by stacking all vectors vi. PCA projects the data into a lower-dimensional space of k components (k ≪ m) as:

$$\mathbf{Z} = \mathbf{V}\mathbf{W}_k \tag{8}$$

Where $\mathbf{W}_k \mathbb{R}^{m\times k}$ contains the top k eigenvectors of the covariance matrix:

$$\mathbf{C} = \frac{1}{N}\mathbf{V}^T\mathbf{V} \tag{9}$$

The final reduced feature vector for each sample is:

$$\mathbf{z}_i = \mathbf{W}_k^T \mathbf{v}_i \in \mathbb{R}^k \tag{10}$$

This reduction step improves computational efficiency and reduces noise, which is critical when working with sparse representations of short social media text.

3.6 Quantum Model Learning

After obtaining reduced feature vectors $\mathbf{z}_i$, multiple quantum learning models are trained under the same conditions to ensure a fair comparison. Each model produces a predicted class label:

$$\hat{y}_i = f_\theta(\mathbf{z}_i), \tag{11}$$

where θ represents model parameters.

3.6.1 Quantum Support Vector Machine (QSVM)

QSVM performs classification using kernel-based separation in a quantum feature space. Given two samples $\mathbf{z}_i$ and $\mathbf{z}_j$, a quantum feature map encodes them as quantum states $|\Phi(\mathbf{z})|$, and the kernel is defined as:

$$K(\mathbf{z}_i, \mathbf{z}_j) = |\langle \Phi(\mathbf{z}_i) | \Phi(\mathbf{z}_j) \rangle|^2 \tag{12}$$

The decision function of QSVM can be written as:

$$f(\mathbf{z}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i \, y_i K(\mathbf{z}_i, \mathbf{z}) + b \right) \tag{13}$$

Where α_i are support vector coefficients, and b is the bias term. This formulation enables non-linear separation of bullying and non-bullying samples in the transformed space.

3.6.2 Quantum Neural Network (QNN)

QNN is implemented using parameterized quantum circuits. For a given input **z**, the circuit prepares a state:

$$|\psi(\mathbf{z}, \theta)\rangle = U(\mathbf{z}, \theta)|0\rangle \tag{14}$$

where $U(\mathbf{z}, \theta)$ is a parameterized unitary transformation. The model prediction is obtained from measurement of an observable M:

$$p(y = 1|\mathbf{z}) = \langle \psi(\mathbf{z}, \theta) | M | \psi(\mathbf{z}, \theta) \rangle \tag{15}$$

The predicted label is given by thresholding:

$$\hat{y} = \begin{cases} 1, & p(y = 1 \mid \mathbf{z}) \geq 0.5, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Model parameters are optimized by minimizing binary cross-entropy:

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{17}$$

where $p_i = p(y = 1|\mathbf{z}_i)$

3.6.3 Quantum Natural Language Processing (QNLP)

QNLP aims to model sentence meaning using structured representations that reflect word composition. The document representation is formed through a composition function:

$$s = g(e_{w_1}, e_{w_2}, \dots, e_{w_L}), \tag{18}$$

Where e_{w_j} denotes the embedding of word w_j and $g(\cdot)$ captures compositional semantics. The composed vector is mapped into a quantum state and classified through measurement-based decision rules. QNLP is evaluated in this work to study whether quantum-based linguistic modelling improves cyberbullying detection performance.

3.6.4 Quantum Singular Value Decomposition (QSVD)

QSVD is included to study decomposition-based quantum learning for text representation. Given the reduced feature matrix **Z**, the singular value decomposition is written as:

$$\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{19}$$

where $\mathbf{\Sigma}$ contains singular values that indicate dominant directions of variance. A compact approximation can be expressed using rank-*r* components:

$$\mathbf{Z}_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T \tag{20}$$

The resulting representations are used to train the classifier and produce the final prediction.

3.7 Hybrid Model (QNN + QSVM)

To improve prediction stability across quantum methods, a hybrid model is proposed by combining QNN and QSVM.

Let $pQNN$ (y = 1 | z) denote the probability predicted by QNN and
$pQSVM$(y = 1|z) denotes the probability obtained from QSVM decision scores.

The hybrid score is computed as:

$$p_{hyb}(y=1 \mid \mathbf{z}) = \lambda\, p_{QNN}(y=1 \mid \mathbf{z}) + (1-\lambda)\, p_{QSVM}(y=1 \mid \mathbf{z}), \tag{21}$$

where $0 \leq \lambda \leq 1$ is a weighting factor. For equal contribution, $\lambda = 0.5$.
The final hybrid decision is

$$\hat{y}_{hyb} = \begin{cases} 1, & p_{hyb}(y=1 \mid \mathbf{z}) \geq 0.5, \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

3.8 Evaluation Metrics

The performance of all models is measured using accuracy, precision, recall, and F1 score, computed from the confusion matrix terms True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{24}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{25}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{26}$$

These metrics provide a balanced evaluation of the detection system, ensuring that the model is assessed not only by overall correctness but also by its reliability in identifying cyberbullying content.

4. Experimental Setup and Results

This section describes the experimental configuration used to evaluate the proposed quantum machine learning framework for cyberbullying detection. All models are tested using the same preprocessing and feature representation pipeline to ensure that performance differences are due to the learning strategy rather than differences in data preparation. The evaluation compares QSVM, QNN, QNLP, QSVD, and the hybrid QNN+QSVM model using standard classification measures.

4.1 Experimental Setup

The experiments follow a unified workflow consisting of text preprocessing, TF– IDF vectorization, and PCA-based dimensionality reduction before applying quantum learning models. This design provides a controlled environment in which each model receives identical input representations.

4.1.1 Text Representation

Each social media post is cleaned and transformed into a TF–IDF feature vector. Since TF–IDF produces a high-dimensional sparse space, PCA is applied to project the features into a reduced set of components. This step improves computational efficiency and helps suppress redundant information. Let the final feature representation of sample xi be denoted as:

$$z_i \in R^k. \tag{27}$$

4.1.2 Model Training and Testing

The reduced feature vectors are then used to train and test four quantum models (QSVM, QNN, QNLP, and QSVD), as well as the hybrid QNN+QSVM ensemble. All models are trained under identical conditions and evaluated on the same test set to ensure fairness and reproducibility.

4.1.3 Evaluation Criteria

To obtain a balanced assessment, the models are evaluated using accuracy, precision, recall, and F1-score. These measures are computed using the confusion matrix statistics TP, TN, FP and FN, as defined in Section 3. The reported values highlight both overall correctness and the ability to reliably detect cyberbullying samples.

4.2 Results and Comparative Analysis

Table 1 presents the experimental performance of the evaluated quantum models using the proposed cyberbullying detection pipeline. All models were trained and tested using the same processing workflow, which includes text cleaning, TF–IDF feature extraction, PCA-based dimensionality reduction, and binary label mapping, where age, ethnicity, gender, religion, and other cyberbullying are grouped as cyberbullying, while not cyberbullying is treated as non-bullying. The experimental outcomes demonstrate that the proposed framework provides strong detection performance, while also revealing clear differences between kernel-driven, decomposition-driven, and circuit-inspired quantum learning approaches.

Table 1. Comparative Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
QSVM	88.15	92.52	92.23	92.38
QNN	84.12	88.34	91.70	89.99
QNLP	83.85	87.56	92.38	89.90
QSVD	88.08	93.53	90.98	92.24
Hybrid (QNN+QSVM)	87.37	92.06	91.68	91.87

As reported in Table 1, QSVM achieved the best overall performance with an accuracy of 88.15%, precision of 92.52%, recall of 92.23%, and an F1-score of 92.38%. The consistently high precision and recall indicate that QSVM maintains an effective balance between limiting false alarms and correctly identifying bullying content. This behaviour is expected from margin-based learning, where the classifier focuses on forming a stable decision boundary in the transformed feature space obtained after TF–IDF and PCA processing.

The QSVD model produced performance very close to QSVM, achieving 88.08% accuracy and an F1-score of 92.24%. Notably, QSVD attained the highest precision of 93.53% among all tested models, which reflects fewer false cyberbullying predictions. This result suggests that the decomposition-based representation can preserve highly discriminative information even after dimensionality reduction, contributing to clean separation of cyberbullying and non-bullying classes.

The hybrid QNN+QSVM model achieved 87.37% accuracy with precision of 92.06%, recall of 91.68%, and an F1-score of 91.87%. Although the hybrid model does not exceed the best standalone accuracy, it demonstrates strong overall stability, with balanced values across all measures. This indicates that combining a circuit-inspired learner with a kernel-based classifier can provide reliable predictions and reduce dependence on a single modelling strategy.

QNN and QNLP obtained competitive results, with accuracies of 84.12% and 83.85%, respectively, and F1-scores close to 90%. The recall obtained by QNLP (92.38%) was the highest among all models, indicating that it is highly sensitive to bullying-related content and is less likely to miss bullying instances. However, this high recall is accompanied by relatively lower precision (87.56%), suggesting an increased number of false positives. Similarly, QNN achieved strong recall (91.70%), but its precision (88.34%) remained lower than the leading methods, indicating that a portion of non-bullying content may be incorrectly flagged as bullying.

Figure 2 illustrates the binary class distribution used for evaluation. The performance comparison across accuracy is visualized in Figure 3, where QSVM and QSVD appear as the best-performing standalone models. Figure 4 further highlights metric-level behavior, showing that QSVM provides consistently high precision and recall, while QSVD offers the strongest precision, and QNLP provides the strongest recall. These trends confirm that the proposed pipeline supports reliable cyberbullying detection and that the most effective models provide balanced behaviour across evaluation measures rather than optimizing accuracy alone.

Fig. 2 Binary class distribution of the cyberbullying dataset after label mapping.

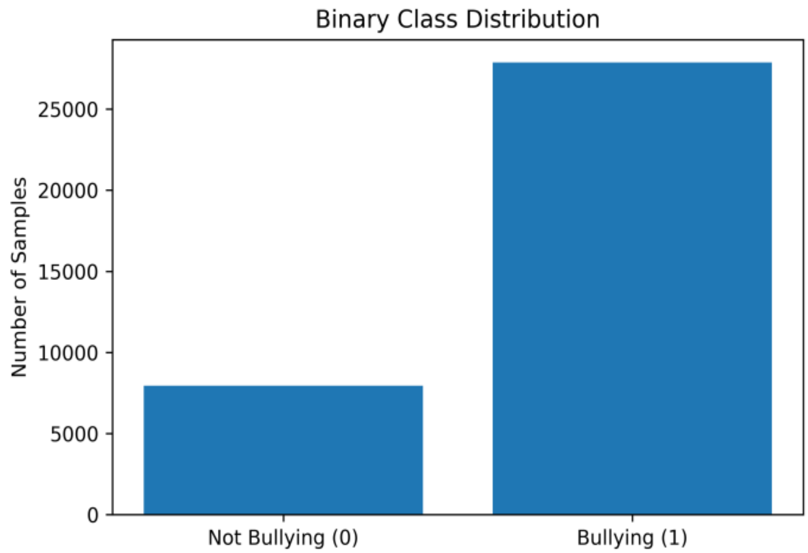


Fig. 3 Accuracy comparison of the evaluated quantum models for cyberbullying detection.

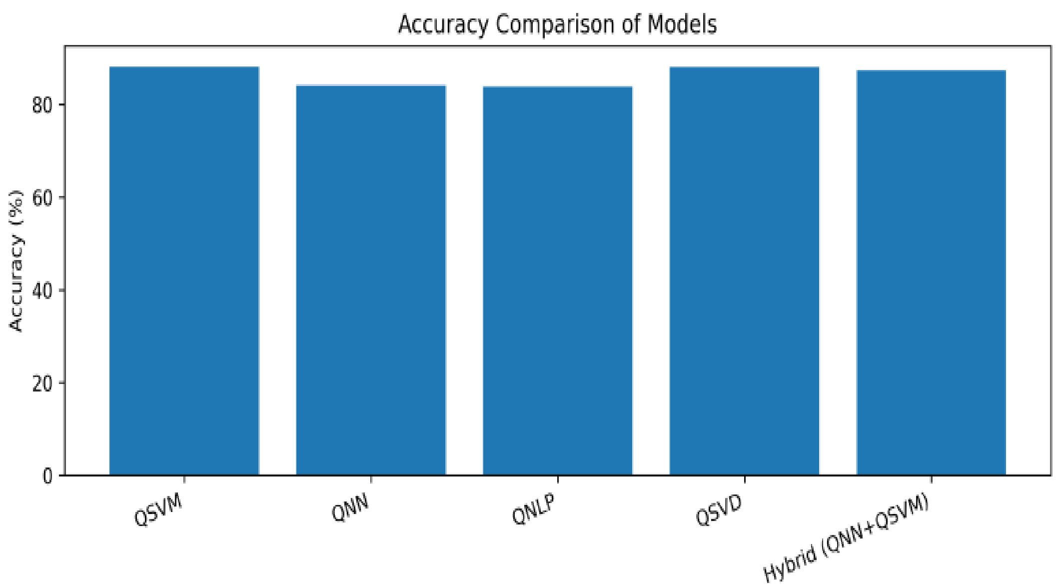
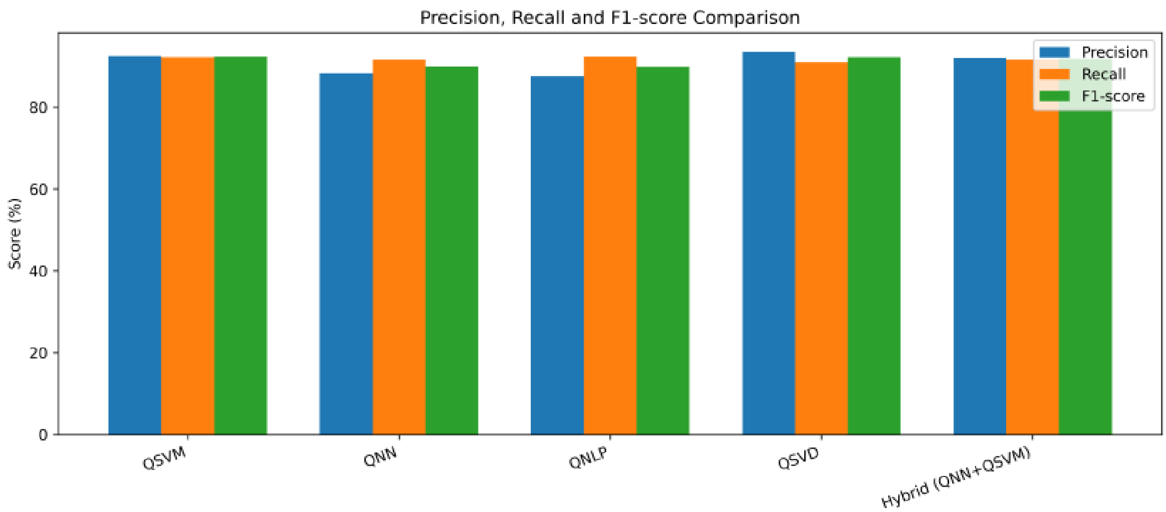


Fig. 4 Precision, recall, and F1-score comparison across the evaluated quantum models.



In addition to metric-level comparisons, confusion matrices provide a detailed view of prediction behaviour. Figures 5–9 show the raw confusion matrices, whereas Figures 10–14 present the normalized confusion matrices. The normalized plots are particularly useful for observing class-wise prediction patterns, confirming that the leading models reduce both false positives and false negatives compared to weaker baselines.

Fig. 5 Confusion matrix of the QSVM classifier.

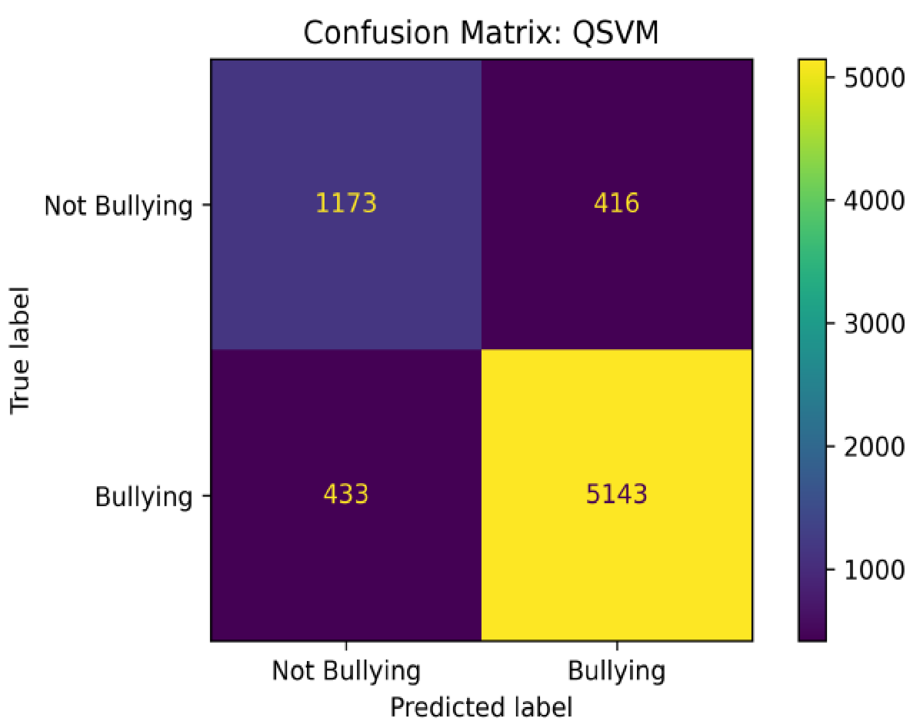


Fig. 6 Confusion matrix of the QNN classifier.

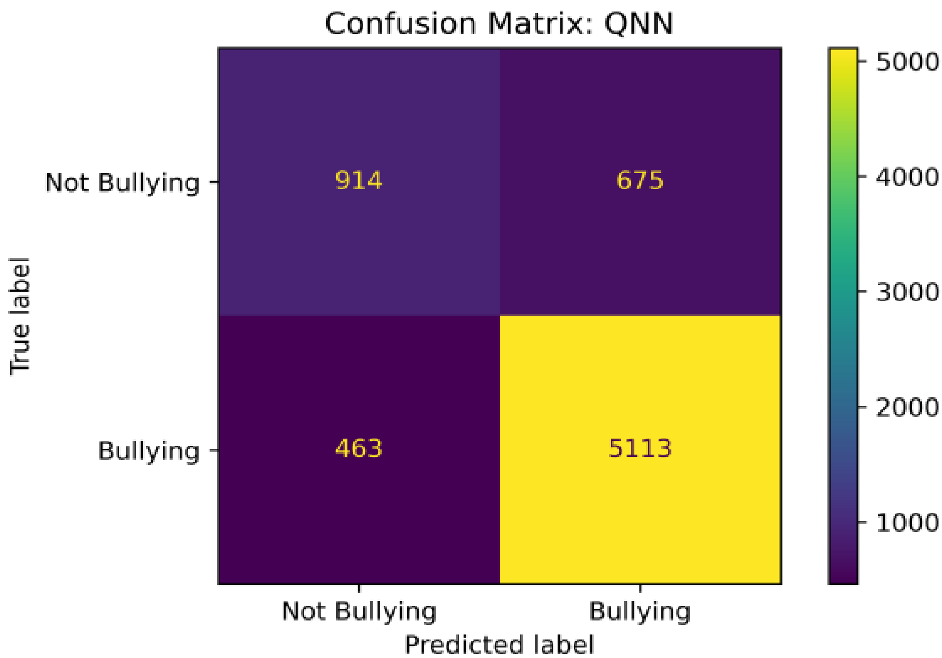


Fig. 7 Confusion matrix of the QNLP classifier.

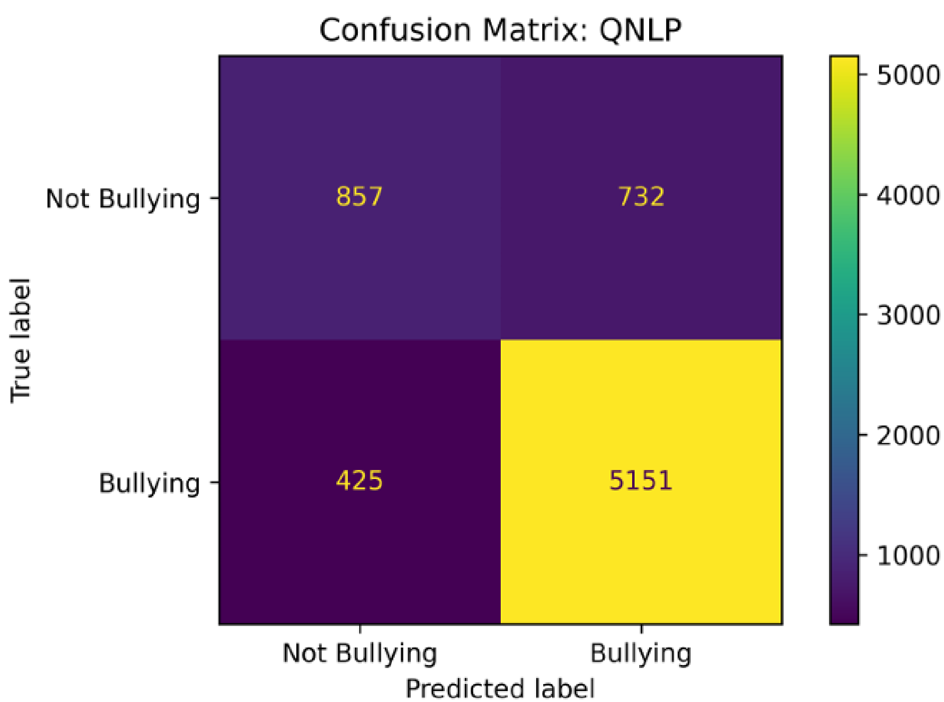


Fig. 8 Confusion matrix of the QSVD classifier.

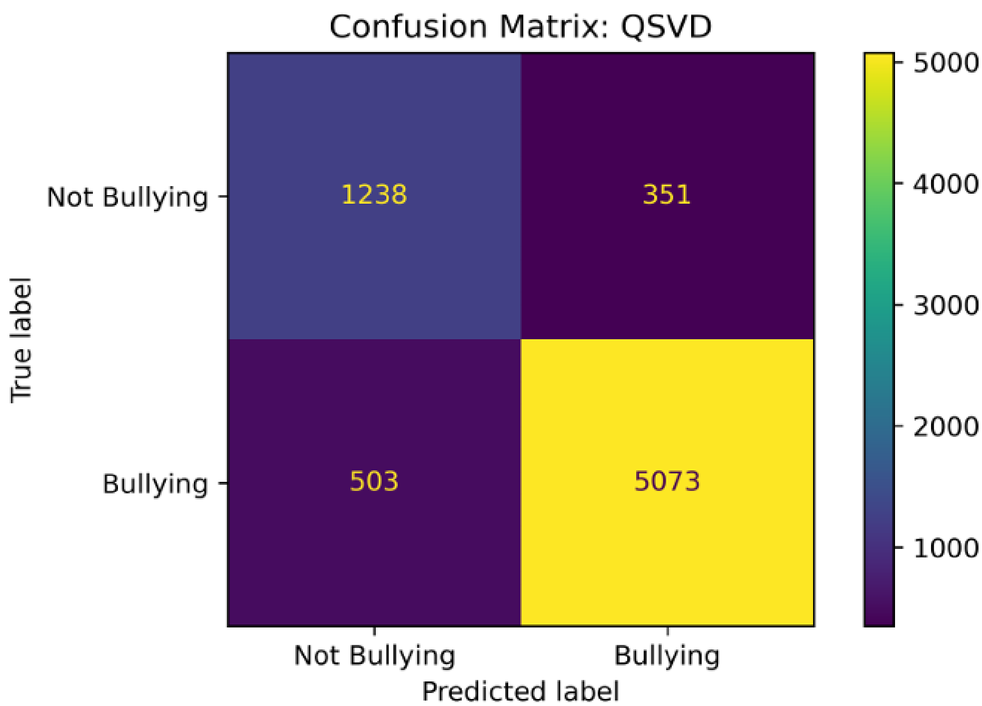


Fig. 9 Confusion matrix of the hybrid QNN+QSVM model.

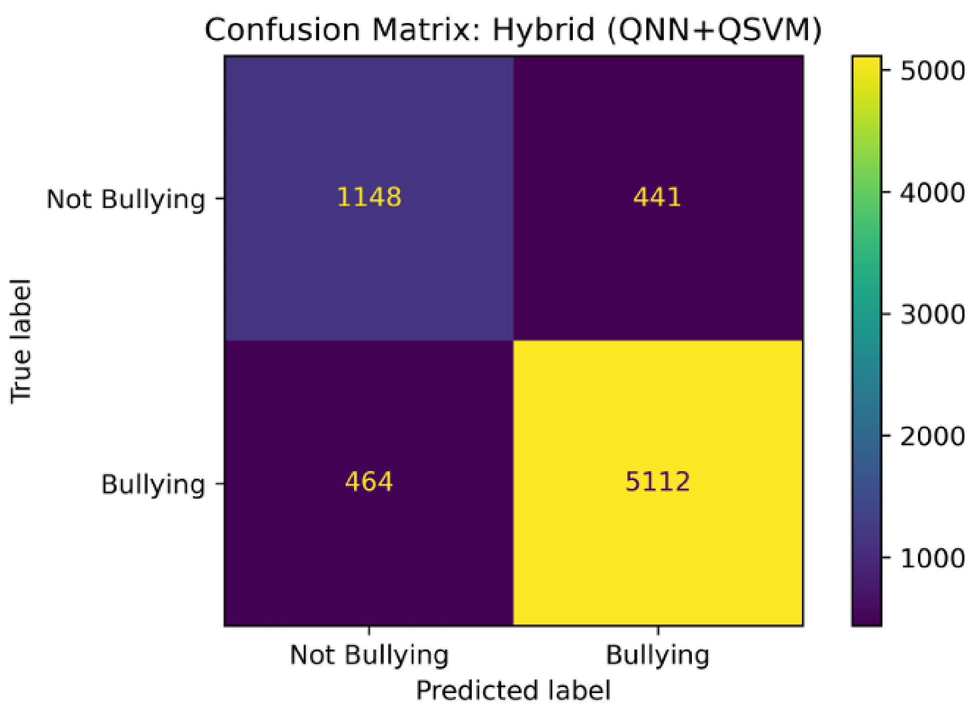


Fig. 10 Normalized confusion matrix of QSVM showing class-wise prediction proportions.

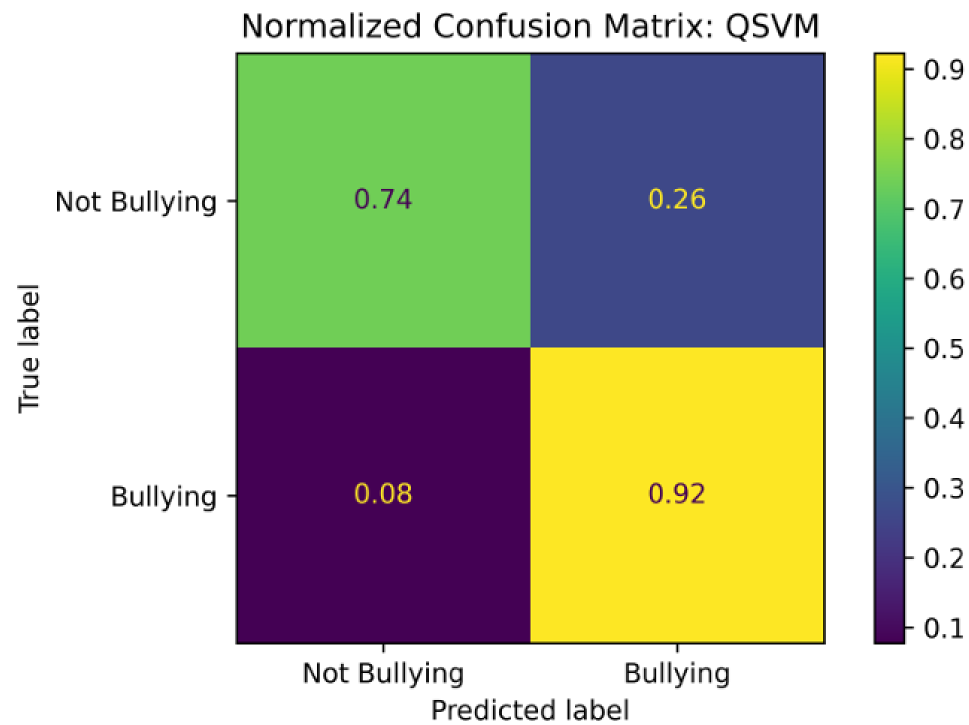


Fig. 11 Normalized confusion matrix of QNN showing class-wise prediction proportions.

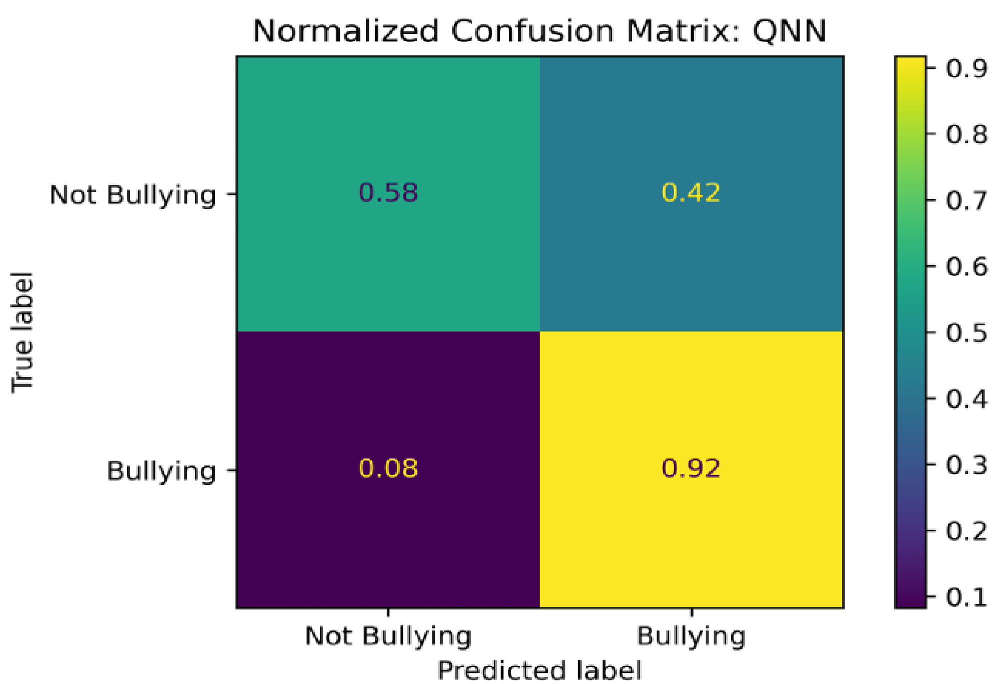


Fig. 12 Normalized confusion matrix of QNLP showing class-wise prediction proportions.

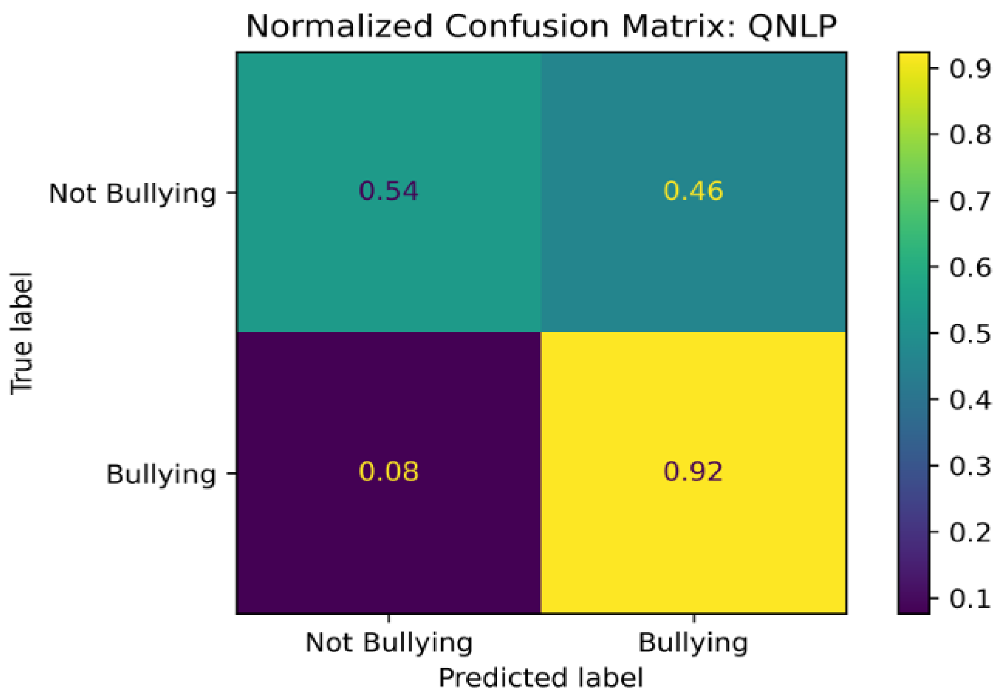


Fig. 13 Normalized confusion matrix of QSVD showing class-wise prediction proportions.

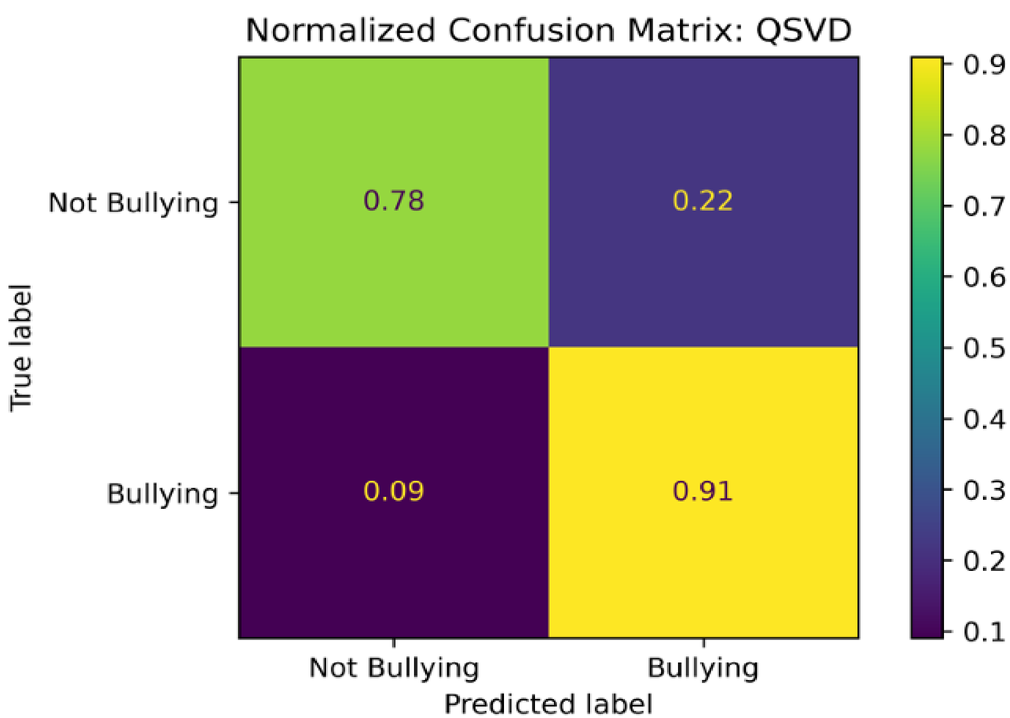
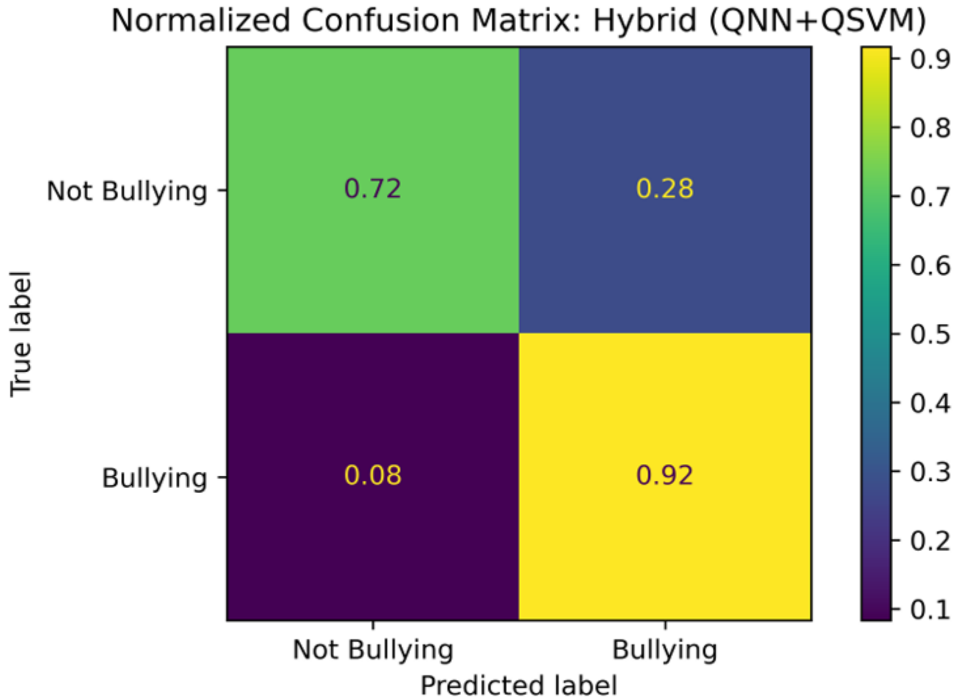


Fig. 14 Normalized confusion matrix of the hybrid QNN+QSVM model showing class-wise prediction proportions.



5. Conclusion

This paper presented a quantum machine learning based framework for cyberbullying detection on social media text using a consistent and reproducible pipeline comprising text preprocessing, TF–IDF feature extraction, and PCA-based dimensionality reduction. Four quantum-inspired learning approaches (QSVM, QNN, QNLP, and QSVD) were implemented and evaluated under identical conditions, along with a hybrid ensemble model that combines QNN and QSVM to examine the benefit of integrating complementary decision mechanisms.

Experimental results demonstrate that the proposed framework can detect cyberbullying content with high reliability. Among the standalone models, QSVM achieved the best overall performance, obtaining 88.15% accuracy with 92.52% precision, 92.23% recall, and an F1-score of 92.38%. QSVD produced comparable accuracy (88.08%) and delivered the highest precision (93.53%), indicating fewer false cyberbullying predictions and confirming the strength of decomposition-based representations in the reduced feature space. The hybrid QNN+QSVM model achieved 87.37% accuracy with a balanced performance profile (F1-score of 91.87%), validating that ensemble integration can maintain stable detection behaviour across evaluation measures. While QNN and QNLP reported slightly lower overall accuracy, both achieved strong recall values, showing their capability to identify bullying instances effectively. Overall, the findings confirm that kernel-based and decomposition-driven quantum models provide strong and consistent performance for cyberbullying detection under the proposed pipeline, while hybrid modelling offers a practical alternative when a balanced precision–recall trade-off is required. Future work will focus on extending the study to larger multilingual datasets, incorporating contextual sentence representations beyond TF–IDF, and investigating optimized quantum circuit designs and hybrid quantum–classical training strategies to further improve generalization and reduce false alarms in real-world deployments.

References

1. Philipo, A. G., Sarwatt, D. S., Ding, J., Daneshmand, M., & Ning, H. (2026). Cyberbullying detection: Exploring datasets, technologies, and approaches on social media platforms. ACM Computing Surveys, 58(7), 1-35.
2. Philipo, A., Ding, J., Sarwatt, D., Mohamed, J., Yusufu, A., Daneshmand, M., & Ning, H. (2026). Sentiment-enhanced cyberbullying detection models on social media platforms. ACM Transactions on the Web, 20(1), 1-26.
3. Hazra, S., Mukherjee, P., Kundu, S., Karmakar, S., Roy, S., & Bhadra, S. (2026, March). Cyber Shield: Cyberbullying Text Detection in Social Media. In 2026 7th International Conference on Computational Intelligence and Networks (CINE) (pp. 1-4). IEEE.
4. Pabla, L. K., Jain, P. K., Patel, P., & Shukla, S. Deep Learning for the Identification of Cyberbullying in Social Networks. Proceedings Copyright, 114, 118.
5. Subhashree, K., and S. Manoj Kumar. "Enhanced quantum long short-term memory neural network based multi-task learning for sentimental analysis and cyberbullying detection." Expert Systems with Applications 282 (2025): 127555.
6. Olaitan, Ololade Funke, et al. "Quantum computing in artificial intelligence: A review of quantum machine learning algorithms." Path of Science 11.5 (2025): 7001-7009.
7. Gokhale, Shreeya Sanjeev, Rishit Ravi Dhote, and Radhakrishnan Delhibabu. "A review of quantum machine learning algorithms, applications, and emerging advantages." Discover Computing 29.1 (2026): 226.
8. Rodriguez-Diaz, Francesc, et al. "A survey of quantum machine learning: Foundations, algorithms, frameworks, data and applications." ACM Computing Surveys 58.4 (2025): 1-35.
9. Syafiq, A. M., Ab Razak, M. F., Abidin, A. F. Z., Mohamad, S., & Kamaruddin, N. K. (2025). Social network approach for cyberbullying detection using machine learning. Journal of Governance and Integrity, 8(1), 874-880
10. Islam, F., Abdullah, A., Emon, M. Z. H., & Nur, F. N. (2025, July). Ensemble Learning for Cyberbullying Detection: A Transformer-Based Approach on English and Bangla Texts. In 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN) (pp. 1-6). IEEE.
11. J. O. Atoum, Cyberbullying detection through sentiment analysis, in Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, Las Vegas, NV, USA, 2020, pp. 292-297.
12. K. Maity, P. Jha, R. Jain, S. Saha, and P. Bhattacharyya, explain thyself bully: Sentiment-aided cyberbullying detection with explanation, in Proceedings of the Document Analysis and Recognition (ICDAR 2023), G. A. Fink, R. Jain, K. Kise, and R. Zanibbi, Eds., Springer Nature Switzerland, Cham, 2023, pp. 132-148.
13. G. Ciaburro, G. Iannace, and V. Puyana-Romero, Sentiment analysis-based method to prevent cyberbullying, in Proceedings of the 2021 International Conference on Wireless Communications, Networking and Applications, Z. Qian, M. A. Jabbar, and X. Li, Eds., Springer Nature Singapore, Singapore, 2022, pp. 721-735.
14. O. Berjawi, R. Khatoun, W. Fahs, and G. Fenza, Sentiment and emotion analysis for cyberbullying detection, in Proceedings of the 2024 11th International Conference on Social Networks Analysis, Management and Security (SNAMS), IEEE, Abu Dhabi, UAE, 2024, pp. 81-86.
15. A. Wirawan, H. D. Cahyono, and Winarno, Easy data augmentation in sentiment analysis of cyberbullying, in 2023 6th International Conference on Information and Communications Technology (ICOIACT), IEEE, Yogyakarta, Indonesia, 2023, pp. 443-447.
16. J. O. Atoum, Cyberbullying detection neural networks using sentiment analysis, in Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, Las Vegas, NV, USA, 2021, pp. 158-164.
17. A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O Brien, A variational eigenvalue solver on a photonic quantum processor, Nature Communications, vol. 5, no. 2, pp. 1-10, 2014.
18. J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New Journal of Physics, vol. 18, no. 2, pp. 1-20, 2016.
19. A. Narayanan and T. Menneer, Quantum artificial neural network architectures and components, Information Sciences, vol. 128, no. 3, pp. 231-255, 2000.
20. Y. Cao, G. G. Guerreschi, and A. Aspuru-Guzik, Quantum neuron: An elementary building block for machine learning on quantum computers, 2017, pp. 1-30.K. Tajeddini, E. Martin and L. Altinay. "The importance of human-related factors on service innovation and performance", *International Journal of Hospitality Management*, 85, Article 102431, 2020.
21. M. Schuld, I. Sinayskiy, and F. Petruccione, Simulating a perceptron on a quantum computer, Physics Letters A, vol. 379, no. 7, pp. 660-663, 2015.
22. B. Ricks and D. Ventura, Training a quantum neural network, in Advances in Neural Information Processing Systems, 2004.
23. M. Panella and G. Martinelli, Neuro fuzzy networks with nonlinear quantum learning, IEEE Transactions on Fuzzy Systems, vol. 17, no. 3, pp. 698-710, 2009.
24. De Vos and Y. Van Rentergem, Synthesis of reversible logic for nano electronic circuits, International Journal of Circuit Theory and Applications, vol. 35, no. 3, 2007.
25. M. Panella and G. Martinelli, Neural networks with quantum architecture and quantum learning, International Journal of Circuit Theory and Applications, vol. 39, no. 1, pp. 61-77, Jan. 2011.
26. R. P. Feynman, Quantum mechanical computers, Foundations of Physics, vol. 16, no. 6, 1986.
27. D. Ventura and T. Martinez, Quantum associative memory, International Journal of Modern Physics B, vol. 17, no. 12, pp. 2447-2472, Jul. 1998.
28. C. A. Trugenberger, Probabilistic quantum memories, Physical Review Letters, vol. 87, no. 6, pp. 67901-1, 67901-4, 2001.
29. C. A. Trugenberger, Quantum pattern recognition, in Invited Talk at the 1st Feynman Festival, University of Maryland,

2002.
30. Bisio, G. Chiribella, G. M. D Ariano, S. Facchini, and P. Perinotti, Optimal quantum learning of a unitary transformation, Physical Review A, vol. 81, no. 3, 2010.
31. S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, Nature Physics, vol. 10, no. 9, pp. 631-633, 2014.
32. W. R. de Oliveira, Quantum RAM-based neural networks, in ESANN 2009 Proceedings, 17th European Symposium on Artificial Neural Networks, 2009, pp. 331-336.
33. J. Silva, T. B. Ludermir, and W. R. Oliveira, Superposition-based learning algorithm, in Proceedings of the 11th Brazilian Symposium on Neural Networks (SBRN), 2010, pp. 1-6.
34. Silva, W. de Oliveira, and T. Ludermir, A weightless neural node based on a probabilistic quantum memory, in Proceedings of the 11th Brazilian Symposium on Neural Networks (SBRN), 2015, pp. 259-264.
35. W. R. de Oliveira, A. J. Silva, T. B. Ludermir, A. Leonel, W. R. Galindo, and J. C. C. Pereira, Quantum logical neural networks, in Proceedings of the 10th Brazilian Symposium on Neural Networks (SBRN), 2008, pp. 147-152.