# Can Variational Quantum Classifiers Learn from Sentence Embeddings? A Controlled Evaluation under NISQ Constraints

**Anonymous ACL submission**


## Abstract

Quantum natural language processing aims to connect linguistic structure and semantic representations with quantum models. Yet many NISQ-era pipelines require sentence embeddings to be compressed to only a few features before quantum encoding. Can variational quantum classifiers (VQCs) still exploit the semantic signal that survives this compression? We test this question on SST-2, a binary sentiment classification benchmark, using PCA-compressed embeddings from BERT, MiniLM, and MPNet. VQCs are compared against classical classifiers trained on exactly the same reduced representations across data sizes, random seeds, and quantum configurations. The result is a sharp mismatch: classical models reach up to 85.6% accuracy, showing that the compressed embeddings retain substantial task signal, while VQCs remain near chance across the tested settings. This rules out a simple "compression destroys the signal" explanation. Instead, the tested VQC pipeline fails to exploit semantic information that remains accessible to classical models. More broadly, our findings show why NLP evaluations of new computational paradigms need matched-input controls that separate representation quality from model-specific benefit.


## 1 Introduction

Quantum natural language processing (QNLP) has attracted growing interest as a way to connect linguistic structure and semantic representations with quantum models (Coecke et al., 2020; Lloyd et al., 2020). A central motivation is that quantum circuits operate in high-dimensional Hilbert spaces, suggesting a possible interface between language representations and quantum feature spaces (Schuld and Killoran, 2019; Havlíček et al., 2019). At the same time, modern NLP relies heavily on dense sentence embeddings produced by transformer models such as BERT (Devlin et al., 2018). These embeddings often contain hundreds of dimensions, while current noisy intermediate-scale quantum (NISQ) devices remain constrained by qubit counts, gate noise, and circuit depth (Preskill, 2018).

This creates a central bottleneck for practical QNLP. Before sentence embeddings can be encoded into many NISQ-compatible circuits, they must be compressed to a small number of features. The key question is therefore not whether quantum classifiers are expressive in principle, but whether they can exploit the semantic signal that remains after compression. This distinction is crucial. If compression destroys the task-relevant signal, no classifier should perform well. If classical models still perform well on the same compressed features while quantum models do not, then the limitation is not simply dimensionality reduction. It lies in how the compressed representation is encoded, transformed, and optimized.

We study this question through a matched-input evaluation of variational quantum classifiers (VQCs) on compressed sentence embeddings. We use embeddings from BERT, MiniLM, and MPNet, reduce them with principal component analysis (PCA), and evaluate them on SST-2 sentiment classification. Crucially, all classical and quantum classifiers are trained on exactly the same PCA-reduced input representations. This design separates two questions that are often conflated: whether compressed sentence embeddings still contain task signal, and whether VQCs can turn that signal into accurate predictions.

Our results reveal a striking pattern. Classical classifiers trained on the same low-dimensional PCA features achieve strong performance, with the best configuration reaching 85.6% accuracy. In contrast, VQCs remain close to chance across encoders, training sizes, and tested quantum configurations. Thus, the semantic signal survives compression, but the tested VQC pipeline does not turn it into

competitive predictions.

Overall, this paper makes three contributions: [1]

- We evaluate NISQ-compatible quantum classifiers in a matched-input setting, separating representation loss from model-specific benefit.
- We show that 8-dimensional PCA embeddings retain task signal for classical models, reaching up to 85.6% accuracy, while VQCs remain close to chance.
- We draw a broader lesson for NLP: new computational paradigms need competitive controls under the same input constraints before model-specific benefits can be claimed.

## 2 Background and Related Work

QNLP has been developed from several directions. One line of work builds on categorical compositional distributional semantics, where grammatical structure is mapped to tensor or quantum circuit representations (Coecke et al., 2020). Another line studies quantum feature maps and parameterized quantum circuits as machine learning models for classical data (Havlíček et al., 2019; Schuld and Killoran, 2019; Cerezo et al., 2021). These approaches motivate the hope that quantum models may provide useful non-linear representations for language tasks.

However, practical QNLP faces a representation bottleneck. Transformer-based NLP models produce high-dimensional embeddings (Vaswani et al., 2017; Devlin et al., 2018), while current NISQ-era circuits can only operate on a small number of qubits (Preskill, 2018). As a result, sentence embeddings must usually be compressed before quantum encoding. This compression step can discard task-relevant information, but it is rarely isolated from the behavior of the quantum classifier itself.

Variational quantum algorithms also face optimization challenges. Parameterized quantum circuits are trained through a hybrid loop in which a classical optimizer updates circuit parameters based on quantum measurements (Cerezo et al., 2021). In deeper or poorly initialized circuits, optimization can suffer from barren plateaus, where gradients vanish exponentially and learning becomes ineffective (McClean et al., 2018). These issues make it necessary to evaluate QNLP pipelines against strong classical baselines under matched input conditions.

Our work treats QNLP as a diagnostic evaluation problem. We do not ask whether quantum models are expressive in principle. Instead, we ask whether a practical VQC pipeline can exploit the semantic signal that remains after sentence embeddings are compressed to NISQ-compatible dimensions.

## 3 Experimental Setup

### 3.1 Task and Sentence Embeddings

We evaluate on SST-2 sentiment classification (Socher et al., 2013), a binary sentence-level task. We use three pretrained encoders to obtain sentence embeddings: BERT (`bert-base-uncased`), MiniLM (`all-MiniLM-L6-v2`), and MPNet (`all-mpnet-base-v2`). BERT and MPNet produce 768-dimensional embeddings, while MiniLM produces 384-dimensional embeddings.

To reflect the computational constraints of hybrid quantum-classical training, we sample training subsets of size $N \in \{20, 50, 100, 200\}$. All experiments are repeated across five random seeds: 42, 123, 777, 1024, and 2026. Additional details on sampling, dimensionality reduction, and implementation are provided in Appendix A.

### 3.2 Dimensionality Reduction

Let $X \in \mathbb{R}^{N \times D}$ denote the sentence embedding matrix, where $D$ is the original embedding dimension. We reduce each embedding to $K$ dimensions using PCA:

$$Z = XW, \quad W \in \mathbb{R}^{D \times K}. \tag{1}$$

For each split and random seed, PCA is fitted only on the training subset and then applied to the corresponding evaluation data. The main experiments use $K = 8$, corresponding to an 8-qubit input representation. This deliberately restrictive setting reflects the compression required by NISQ-compatible quantum models. We additionally evaluate $K \in \{4, 8, 10\}$.

### 3.3 Quantum Classifier

The quantum classifier encodes the reduced vector $z_i \in \mathbb{R}^K$ with a ZZ feature map,

$$|\Phi(z_i)\rangle = U_\Phi(z_i)\,|0\rangle^{\otimes K}, \tag{2}$$

followed by an `EfficientSU2` variational ansatz:

$$|\psi(z_i, \theta)\rangle = W(\theta) U_\Phi(z_i)\,|0\rangle^{\otimes K}. \tag{3}$$

[1] Our code and reproducibility instructions are available at https://anonymous.4open.science/r/DiagnosticBenchmarkVQC/

Predictions are obtained from the expectation value of a Pauli observable. We evaluate shallow and deeper ansatz configurations, linear and full entanglement patterns, COBYLA and SPSA optimizers, finite-shot simulation with 1024 shots, and a depolarizing noise model. A more detailed circuit specification is given in Appendix B.

### 3.4 Classical Baselines

To isolate the contribution of the quantum component, we compare the VQC against classical classifiers trained on exactly the same PCA-reduced representations. The baselines are logistic regression, linear SVM, RBF SVM, multilayer perceptron (MLP), and random forest. This matched-input comparison is essential: if classical models succeed on the same compressed features, then the compressed representation cannot be dismissed as unusable semantic noise.

## 4 Results

### 4.1 Matched-Input Controls Reveal Surviving Task Signal

Table 1 summarizes the main results. For each encoder and training size, we report the best classical baseline on the same PCA-reduced features, the corresponding VQC accuracy, and the resulting gap. Full per-encoder results for all classical baselines are provided in Appendix C (Tables 2, 3, and 4).

The matched-input comparison shows a clear pattern. PCA-compressed embeddings are not simply uninformative: classical models trained on the same 8-dimensional inputs achieve strong performance, especially for BERT and MPNet. MPNet reaches 85.6% accuracy with a linear SVM at $N = 200$, while the corresponding VQC obtains 52.3%. BERT shows a similar gap, with classical models around 81–82% accuracy and VQC performance close to chance. MiniLM is weaker overall, but even there the best classical baseline reaches 70.6%, substantially above the VQC.

During preliminary hyperparameter exploration, tailored shallow configurations allowed the VQC to reach a peak accuracy of 60%. However, this localized success proved to be unstable and failed to generalize to broader data distributions. To further investigate this limitation, we attempted to optimize the pipeline by testing Quantum Kernel methods (QSVC) and increasing the feature map complexity. Counterintuitively, these additions led to sharp performance drops, indicating that moving beyond a minimal setup inflates the parameter space. Instead of improving semantic representations, adding complexity creates a highly non-convex optimization landscape that gradient-free optimizers (like COBYLA and SPSA) struggle to navigate effectively, rendering the model overly sensitive to initialization and noise

### 4.2 Better Embeddings Help Classical Models, Not VQCs

The encoder comparison sharpens the result. MPNet provides the strongest compressed representations, followed by BERT and MiniLM. This improves the classical baselines, but it does not translate into improved VQC performance. The gap is therefore largest precisely in the settings where the compressed representations are most useful for classical models.

### 4.3 VQC Ablations Do Not Close the Gap

We further test whether the VQC failure can be explained by circuit size, ansatz depth, optimizer choice, or noise. Increasing the number of qubits from 4 to 8 and 10 increases the dimensionality of the input representation, but does not yield a reliable improvement in VQC accuracy. Similarly, increasing the depth of the `EfficientSU2` ansatz from one to three repetitions expands the parameter space but does not improve generalization. The optimizer and depth ablation is reported in Appendix D (Table 5); additional qubit, entanglement, and noise results are summarized there.

Optimizer choice changes the failure mode but not the conclusion. COBYLA is comparatively stable but slow, frequently exhausting the function-evaluation budget. SPSA can increase training accuracy but does not improve test accuracy: in the additional BERT experiment with $N = 300$, SPSA reaches 59.0% training accuracy but only 44.7% test accuracy. Finite-shot simulation with 1024 shots and depolarizing noise also fails to close the gap. Overall, the tested VQC pipeline remains non-competitive under matched-input controls.

## 5 Discussion

The main implication of our results is that PCA compression is not enough to explain the VQC failure. Reducing sentence embeddings from hundreds

| Encoder | Train size | Best classical model | Classical acc. | VQC acc. | Gap |
|---|---|---|---|---|---|
| BERT | 20 | RBF SVM | 77.9 | 51.6 | +26.3 |
| | 50 | RBF SVM | 81.5 | 50.7 | +30.8 |
| | 100 | Logistic Reg. | 81.4 | 53.5 | +27.9 |
| | 200 | Random Forest | 81.9 | 51.1 | +30.8 |
| MPNet | 20 | RBF SVM | 71.5 | 50.1 | +21.4 |
| | 50 | MLP | 76.0 | 53.9 | +22.1 |
| | 100 | Linear SVM | 82.7 | 51.8 | +30.9 |
| | 200 | Linear SVM | 85.6 | 52.3 | +33.3 |
| MiniLM | 20 | MLP | 57.6 | 49.7 | +7.9 |
| | 50 | RBF SVM | 62.4 | 50.3 | +12.1 |
| | 100 | RBF SVM | 67.0 | 48.4 | +18.6 |
| | 200 | Linear SVM | 70.6 | 51.1 | +19.5 |

Table 1: Main results on SST-2 using 8-dimensional PCA-compressed sentence embeddings. Accuracies and gaps are reported in percentage points. Classical baselines and VQCs are trained on the same reduced features.

of dimensions to only a few features is clearly a severe constraint. Yet the matched-input comparison shows that these reduced representations still support strong linear and kernel-based classifiers. The bottleneck therefore shifts: the problem is not the absence of semantic signal, but the inability of the tested quantum pipeline to turn this signal into accurate predictions.

This points to a more specific failure mode at the interface between representation and model. PCA preserves high-variance directions, and the success of the classical baselines suggests that these directions remain predictive for sentiment classification. Quantum feature maps can in principle induce expressive non-linear spaces (Schuld and Killoran, 2019; Havlíček et al., 2019), but this expressivity is useful only if the induced geometry matches the structure of the data. In our setting, the ZZ feature map and variational ansatz do not appear to improve the compressed representation; instead, they introduce a harder trainable transformation on top of a structure that simpler models can already exploit.

The ablations support this interpretation. Increasing circuit depth, changing the optimizer, or adding qubits does not close the gap to the classical baselines. This is consistent with the known sensitivity of variational circuits to ansatz choice, initialization, and optimization pathologies such as barren plateaus (McClean et al., 2018; Cerezo et al., 2021). The issue is therefore best understood as an interaction between encoding, inductive bias, and trainability, rather than as a lack of information in the compressed embeddings alone.

For the community, the broader lesson is methodological. New computational paradigms are often motivated by architectural expressivity: richer feature spaces, alternative inductive biases, or theoretically powerful transformations. Our results show why this is insufficient. A model can map language representations into a mathematically rich space and still fail to exploit semantic structure that is accessible to simple baselines. Matched-input controls are therefore necessary for deciding whether a new architecture contributes anything beyond the representation it receives. Our findings do not rule out future quantum advantages for language processing, but they shift the burden of proof: such advantages should be demonstrated under controls that separate representation quality from model-specific benefit.

## 6 Conclusion

We presented a controlled evaluation of variational quantum classifiers on PCA-compressed sentence embeddings. Across three encoders, multiple training sizes, several quantum configurations, and matched classical baselines, we find that classical models can exploit the compressed semantic representations while VQCs remain close to chance. This shows that the failure of the tested quantum NLP pipeline cannot be attributed to dimensionality reduction alone.

Future work on QNLP should therefore prioritize diagnostic evaluations that test whether quantum models can exploit semantic signal under matched classical controls before making broader claims about quantum advantage for language representations.

## Limitations

This study focuses on one sentence-level classification task, SST-2, and three transformer-based encoders. The conclusions may not transfer directly to other NLP tasks, such as natural language inference, retrieval, or structured prediction. We also evaluate a specific family of quantum models: VQCs based on ZZ feature-map encodings and `EfficientSU2` ansatzes. Other quantum approaches, such as quantum kernels, data re-uploading architectures, or amplitude encoding, may behave differently.

Our experiments are constrained to low-qubit settings and mostly rely on simulation rather than execution on physical quantum hardware. This is appropriate for controlled diagnosis, but it does not capture all device-specific effects of real hardware. Conversely, simulation also limits the feasible scale of the experiments, especially for larger qubit counts and extensive hyperparameter search.

Finally, our evaluation focuses on whether VQCs can exploit PCA-compressed sentence embeddings. PCA is a simple linear compression method and may not preserve all task-relevant semantic structure. Future work should test whether non-linear compression methods or quantum-native encodings improve the transfer of semantic signal into quantum circuits.

## References


M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and 1 others. 2021. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2020. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. 2019. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567:209–212.

Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. 2020. Quantum embeddings for machine learning. *arXiv preprint arXiv:2001.03622*.

Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. 2018. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812.

John Preskill. 2018. Quantum computing in the nisq era and beyond. *Quantum*, 2:79.

Maria Schuld and Nathan Killoran. 2019. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642.

Ashish Vaswani, Niki Shazeer, Noam Sus e Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.

## A Experimental Details

All models are evaluated on SST-2 sentiment classification. Sentence embeddings are extracted using three pretrained encoders: `bert-base-uncased`, `all-mpnet-base-v2`, and `all-MiniLM-L6-v2`. BERT and MPNet produce 768-dimensional embeddings, while MiniLM produces 384-dimensional embeddings.

For each encoder, we sample training subsets of size $N \in \{20, 50, 100, 200\}$. All experiments are repeated across five random seeds: 42, 123, 777, 1024, and 2026. The same reduced representations are used for both the classical baselines and the variational quantum classifier (VQC), ensuring that performance differences are not caused by different input features.

The main experiments reduce sentence embeddings to $K = 8$ dimensions using principal component analysis (PCA), corresponding to an 8-qubit input representation. We additionally evaluate qubit settings $K \in \{4, 8, 10\}$. Accuracy is reported as the mean over seeds. Classical baselines include logistic regression, linear SVM, RBF SVM, multilayer perceptron (MLP), and random forest.

## B Quantum Circuit Specification

This section provides the mathematical specification of the VQC used in the experiments. Let $z_i \in \mathbb{R}^K$ denote the PCA-reduced sentence embedding of input sentence $i$. The quantum model first maps $z_i$ into a quantum state using a ZZ feature map:

$$|\Phi(z_i)\rangle = U_\Phi(z_i)\,|0\rangle^{\otimes K}\,. \tag{4}$$

The feature-map unitary is composed of layers of Hadamard gates and data-dependent phase rotations. In the standard ZZ feature-map construction, this can be written as:

$$U_\Phi(z_i) = \prod_{l=1}^{d_{\text{feat}}} \left( \exp\left[ i \sum_{j=1}^{K} z_{i,j} Z_j + i \sum_{j<k} (\pi - z_{i,j})(\pi - z_{i,k}) Z_j Z_k \right] H^{\otimes K} \right). \tag{5}$$

After data encoding, we apply a parameterized variational ansatz $W(\theta)$ based on `EfficientSU2`. The resulting state is:

$$|\psi(z_i, \theta)\rangle = W(\theta) U_\Phi(z_i)\,|0\rangle^{\otimes K}\,. \tag{6}$$

The ansatz consists of layers of single-qubit rotations and entangling gates:

$$W(\theta) = \prod_{l=1}^{d_{\text{ansatz}}} \left( U_{\text{ent}} \prod_{j=1}^{K} R_Y(\theta^Y_{j,l}) R_Z(\theta^Z_{j,l}) \right), \tag{7}$$

where $U_{\text{ent}}$ denotes the entanglement layer. We evaluate both linear and full entanglement patterns.

Classification is performed by measuring an observable $O$, such as a Pauli-$Z$ observable, and mapping the expectation value to a binary prediction:

$$\hat{y}_i = \operatorname{sgn}\left( \langle \psi(z_i, \theta) |\, O \,| \psi(z_i, \theta) \rangle \right). \tag{8}$$

The parameters $\theta$ are optimized using gradient-free classical optimizers, including COBYLA and SPSA. The main paper reports the empirical finding that this VQC pipeline remains close to chance despite the presence of usable task signal in the compressed representations.

## C Full Results by Encoder

Tables 2, 3, and 4 report the full classical and quantum results for each encoder. These tables expand the compressed summary shown in the main paper.

Table 2: BERT performance. Mean accuracy over five seeds.

| Train | LogReg | LinSVM | RBFSVM | MLP | RF | VQC |
|---|---|---|---|---|---|---|
| 20 | 0.771 | 0.752 | 0.779 | 0.742 | 0.761 | 0.516 |
| 50 | 0.799 | 0.789 | 0.815 | 0.770 | 0.787 | 0.507 |
| 100 | 0.814 | 0.803 | 0.809 | 0.780 | 0.812 | 0.535 |
| 200 | 0.811 | 0.804 | 0.814 | 0.773 | 0.819 | 0.511 |

Table 3: MPNet performance. Mean accuracy over five seeds.

| Train | LogReg | LinSVM | RBFSVM | MLP | RF | VQC |
|---|---|---|---|---|---|---|
| 20 | 0.532 | 0.658 | 0.715 | 0.703 | 0.640 | 0.501 |
| 50 | 0.711 | 0.730 | 0.754 | 0.760 | 0.748 | 0.539 |
| 100 | 0.812 | 0.827 | 0.815 | 0.772 | 0.796 | 0.518 |
| 200 | 0.842 | 0.856 | 0.853 | 0.838 | 0.837 | 0.523 |

Table 4: MiniLM performance. Mean accuracy over five seeds.

| Train | LogReg | LinSVM | RBFSVM | MLP | RF | VQC |
|---|---|---|---|---|---|---|
| 20 | 0.506 | 0.505 | 0.537 | 0.576 | 0.542 | 0.497 |
| 50 | 0.601 | 0.607 | 0.624 | 0.613 | 0.595 | 0.503 |
| 100 | 0.657 | 0.660 | 0.670 | 0.598 | 0.646 | 0.484 |
| 200 | 0.698 | 0.706 | 0.678 | 0.643 | 0.687 | 0.511 |

## D Additional VQC Ablations

We further evaluate whether VQC performance is limited by ansatz depth or optimizer choice. Table 5 reports an additional ablation using BERT embeddings with $N = 300$ and $K = 8$.

Table 5: VQC optimizer and ansatz-depth ablation using BERT embeddings with $N = 300$ and $K = 8$.

| Optimizer | Ansatz Depth | Max Iter. | Train Acc. | Test Acc. |
|---|---|---|---|---|
| COBYLA | Deeper (`reps=3`) | 100 | 0.5467 | 0.5167 |
| COBYLA | Shallow (`reps=1`) | 100 | 0.5833 | 0.5400 |
| SPSA | Shallow (`reps=1`) | 50 | 0.5900 | 0.4467 |

The shallow COBYLA configuration performs slightly better than the deeper COBYLA configuration, while SPSA increases training accuracy but reduces test accuracy. These results support the main-paper conclusion that varying optimizer and depth changes the failure mode but does not close the gap to classical baselines.

We also evaluate different qubit settings ($K \in \{4, 8, 10\}$), entanglement patterns, finite-shot simulation, and depolarizing noise. The qualitative pattern is consistent with the main results: increasing the number of qubits or changing the entanglement pattern does not produce a reliable improvement in VQC accuracy, and noise-aware simulation further weakens performance. These findings suggest that the tested VQC pipeline is not limited by a single hyperparameter choice alone.

## E Reproducibility and Artifact Notes

The experiments use existing public artifacts: SST-2 for sentiment classification, pretrained sentence encoders for embedding extraction, classical classifiers, and quantum-circuit simulation tools. For reproducibility, the following details should be reported with the released code or configuration files:

- exact versions of the sentence-transformer and transformer libraries;
- exact Qiskit version and simulator backend;
- scikit-learn version for PCA and classical baselines;
- train, validation, and test split construction;
- whether PCA is fitted only on the training subset or on a larger split;
- preprocessing and normalization applied before PCA;
- all classical baseline hyperparameters;
- all VQC circuit parameters, optimizer settings, and random initializations;
- number of shots for shot-based simulation;
- noise-model parameters for noise-aware simulation;
- hardware used for simulation and approximate runtime.

The main paper is self-contained and reports the central empirical finding. The appendix provides additional implementation details and full results to facilitate replication.

## F Libraries used

| Library / Package | Primary Usage |
|---|---|
| Quantum Computing Framework | |
| `qiskit` | Circuit construction & primitives |
| `qiskit-aer` | Statevector & noisy simulation |
| `qiskit-algorithms` | Classical optimizers (COBYLA, SPSA) |
| `qiskit-machine-learning` | VQC, FidelityQuantumKernel |
| Machine Learning & NLP | |
| `scikit-learn` | PCA, classical baselines, metrics |
| `sentence-transformers` | Pre-trained encoder models |
| `datasets` | Hugging Face SST-2 data loading |
| Data Processing & Utilities | |
| `numpy` | Array and matrix operations |
| `pandas` | Results aggregation and CSV export |
| `joblib` | Quantum kernel matrix caching |
| `tqdm` | Iteration progress tracking |

Table 6: List of Python software packages and libraries utilized for data processing, classical machine learning baselines, and quantum circuit simulation.

## G Disclosure of AI Assistance

Generative Artificial Intelligence (Gemini 3 fast) was used to debug and check for inconsistencies within the explanations, as well as for advanced grammar check. AI assistance served only as a technical support role and had no influence on the results presented.