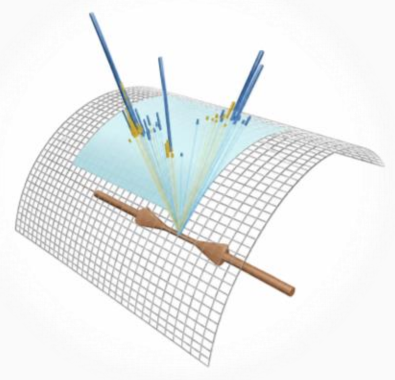
Variational Quantum Classification Heads for Particle Transformer-Based in Jet Tagging

Nadia Ahmed Sharna*
Bursa Technical University, Turkiye

Dr. Izzet F. Senturk
Bursa Technical University, Turkiye

Can a small Variational Quantum Circuit act as a useful classification head on top of Particle Transformer representations?



Main distinction to preserve:

Classifier-head performance vs. information lost by compressing 128 ParT features to only 4–8 dimensions/qubit

Quantum Variational Circuit

- Quantum gates have adjustable parameters

- Initial State / Data Encoding

- The Ansatz (quantum circuit architecture containing tunable gates)

- Measurement & Cost Function

- Classical Optimizers

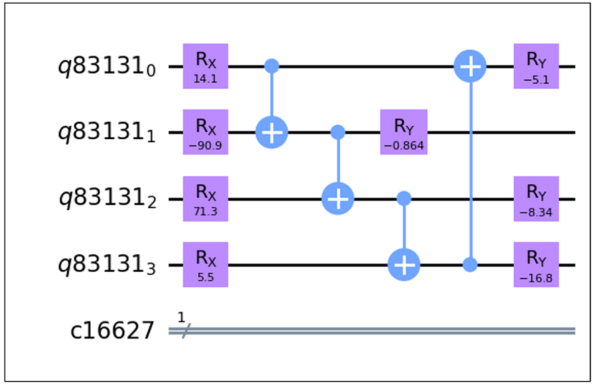


Fig 1.1 Typical variation circuit diagram

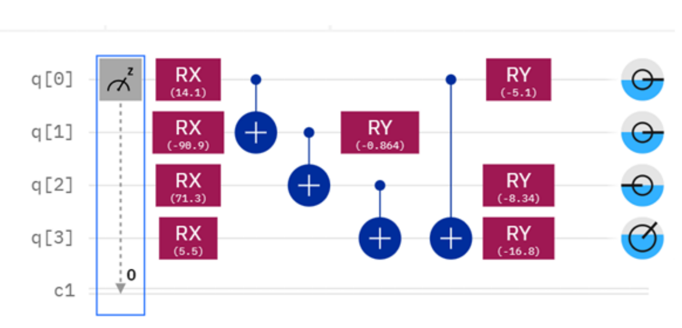


Fig 1.2 The circuit diagram generated in IBM Quantum Composer

Why JetTagging?

Question for every reconstructed jet: what produced this spray of particles?

QCD

background-like jets

Hbb / Wqq / Tbqq

selected signal-like categories

- Jets encode the initiating particle or process.
- Substructure carries correlations among constituents
- Collider searches require robust signal–background separation

JETCLASS DATASET

JetClass is a new large-scale dataset that consists of 100M jets. The dataset contains 10 classes of jets.

Tasks	
Binary	**Multiclass**
QCD vs Hbb	QCD Hbb Wqq Tbqq

CONTROLLED EVENT SELECTION

Constructed a fixed master pool using the global reproducibility seed.

We assigned fixed train, validation, test roles before downstream model comparisons

From a Jet to Particle-Level Model Inputs

Each jet is represented by its constituent particles:

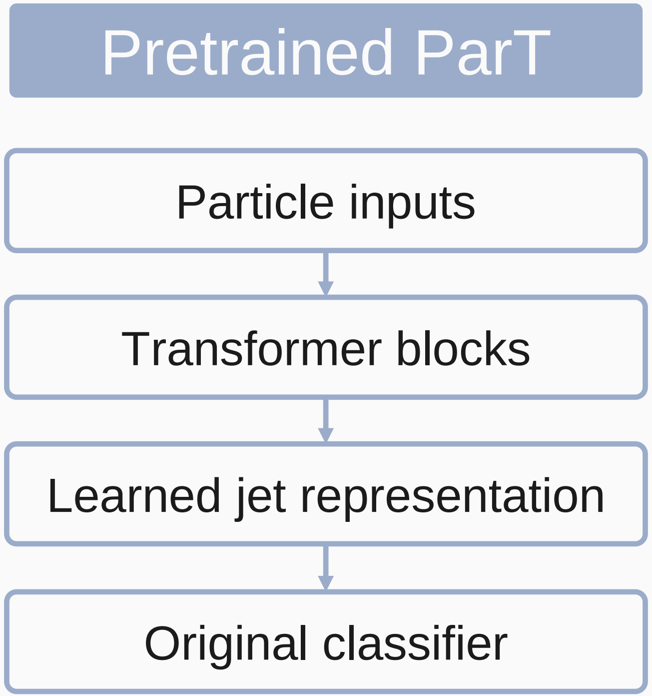
- 17 particle features
- Particle four-vectors:

$$(p_x, p_y, p_z, E)$$

- Constituent mask distinguishes real particles from padding
- Jets are padded/truncated to 128 constituents

Particle Transformer as a Frozen Feature Extractor

- Load pretrained Particle Transformer (ParT)
- Original pretrained weights retained
- Original classification layer removed
- Replace final classifier with Identity
- Set model to evaluation mode
- Disable gradients for all ParT parameters



Frozen Feature Extraction

Each selected jet is passed through frozen ParT:

$$\text{Jet constituents} \rightarrow \text{Frozen ParT} \rightarrow \mathbf{z} \in \mathbb{R}^{128}$$

- Output checked to be **128-dimensional**
- Representations extracted once
- Saved together with:
    - labels
    - data split
    - event indices
- Downstream heads train on these stored representations

Two Representation Branches

COMPRESSED REPRESENTATION

- Start from 128-D ParT vector
- Fit PCA on TRAINING data only
- 128-D → d-dimensional bottleneck
- $d \in \{4,5,6,7,8\}$
- 99th-percentile angle scaling
- Clip to [-1,1]
- Multiply by π
- Final range $\approx [-\pi, \pi]$

UNCOMPRESSED CONTROL

- Start from original 128-D vector
- Standardize using TRAINING data only
- Keep all 128 features
- Purpose: measure how much performance is lost due to compression

Variational Quantum Classification Head

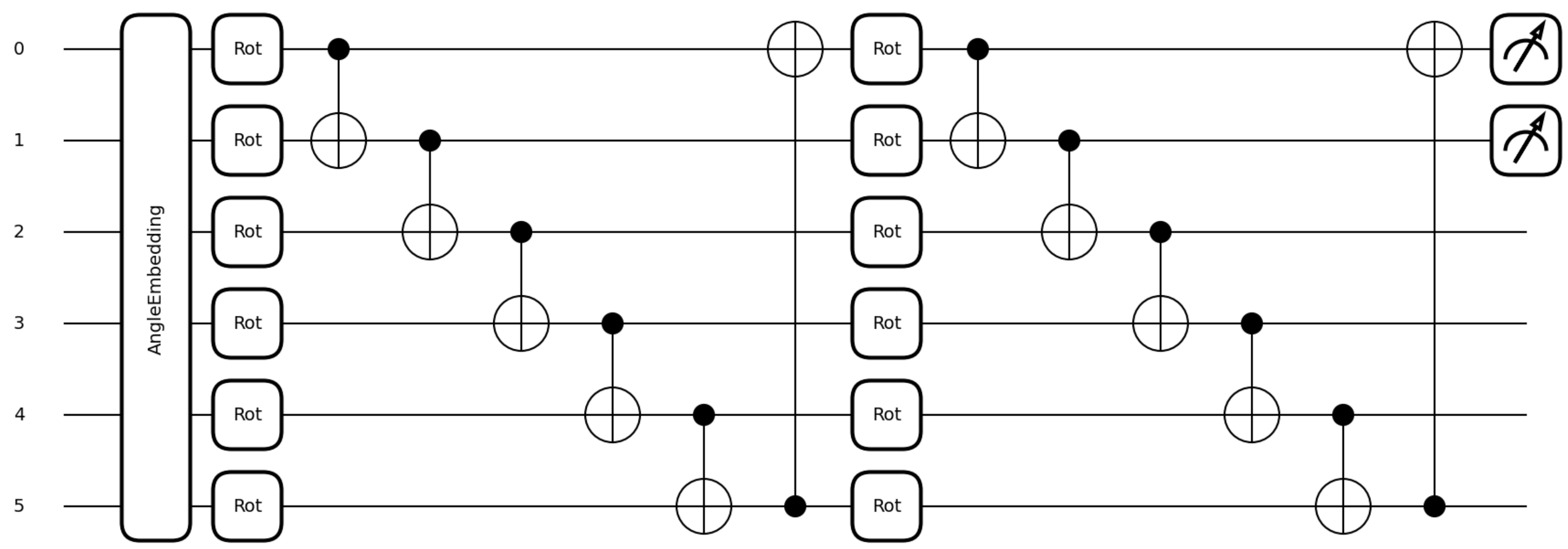


Fig 2.1 Core variational circuit diagram

Quantum Model Configuration

Baseline

- 6 qubits
- Circuit depth = 2
- Ring entanglement
- Analytic statevector simulation
- PennyLane default.qubit
- shots=None
- Backpropagation through the simulator

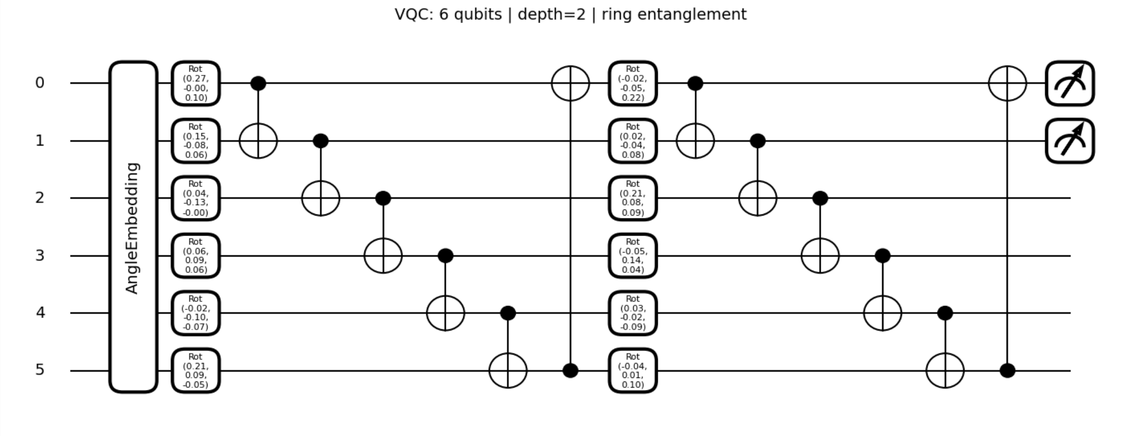


Fig 2.2 Our actual 6-qubit / two-layer circuit

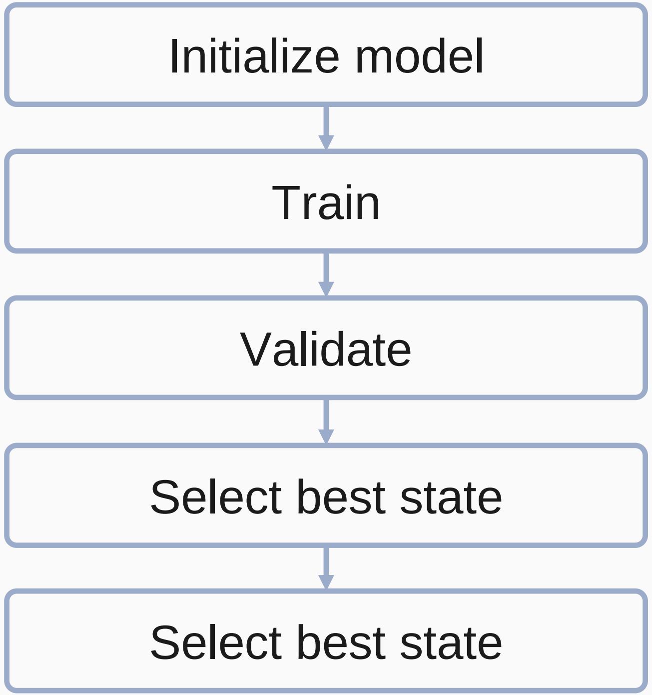
Trainable parameters

Each qubit has 3 trainable rotation parameters per layer:

$$N\_\theta=3nL$$

Model Training Procedure

- Cross-entropy classification loss
- Adam optimizer
- Mini-batch training
- Validation loss monitored after each epoch
- Early stopping prevents unnecessary training
- Best validation model restored before test evaluation



Evaluation Metrics

Classification

- Accuracy
- ROC-AUC
- Multiclass macro one-vs-rest ROC-AUC

Probability quality

- Reliability diagrams
- Expected Calibration Error
- Brier score

Collider-oriented discrimination

Background rejection at:

- 50% signal efficiency
- 80% signal efficiency

Robustness and complexity

- Stability across seeds
- Trainable parameter count
- Training/inference time

Results

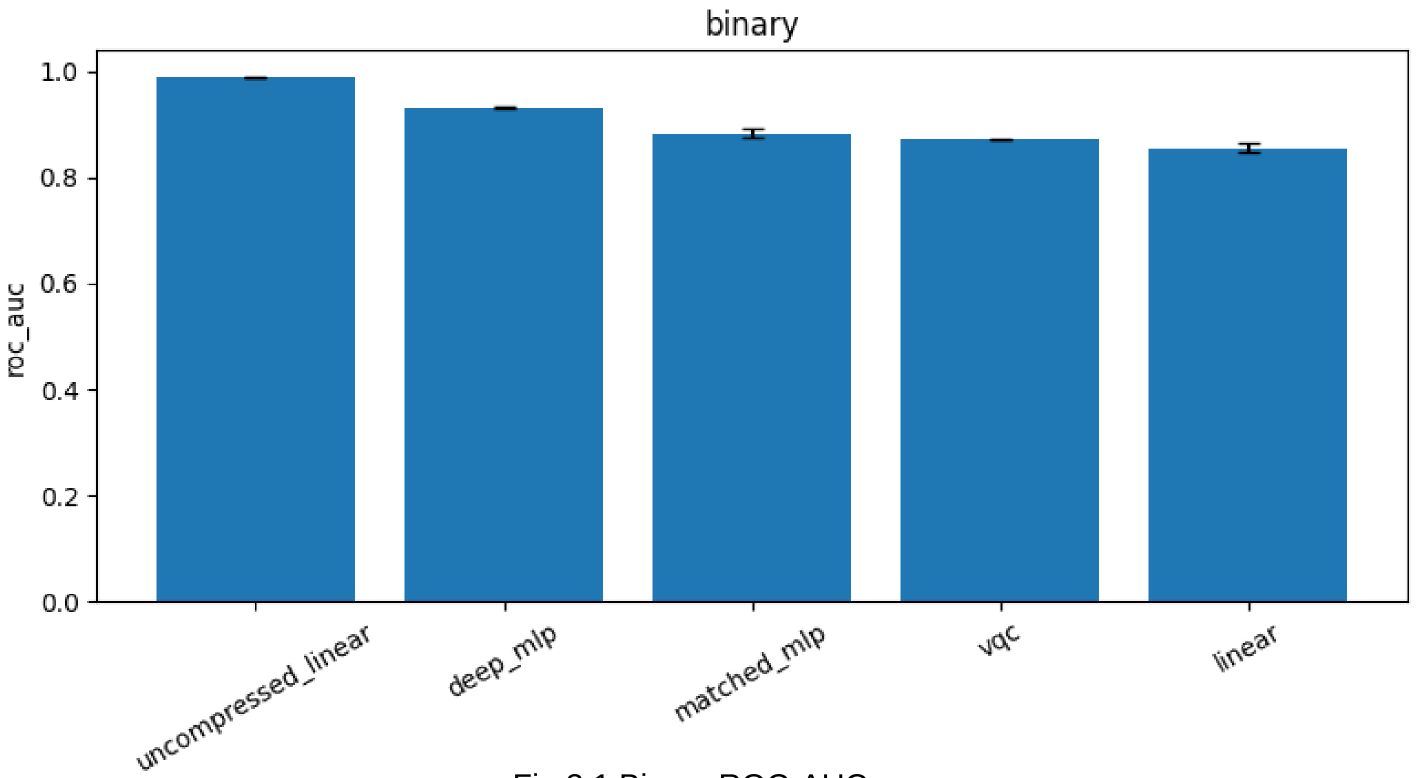


Fig 3.1 Binary ROC-AUC

Results

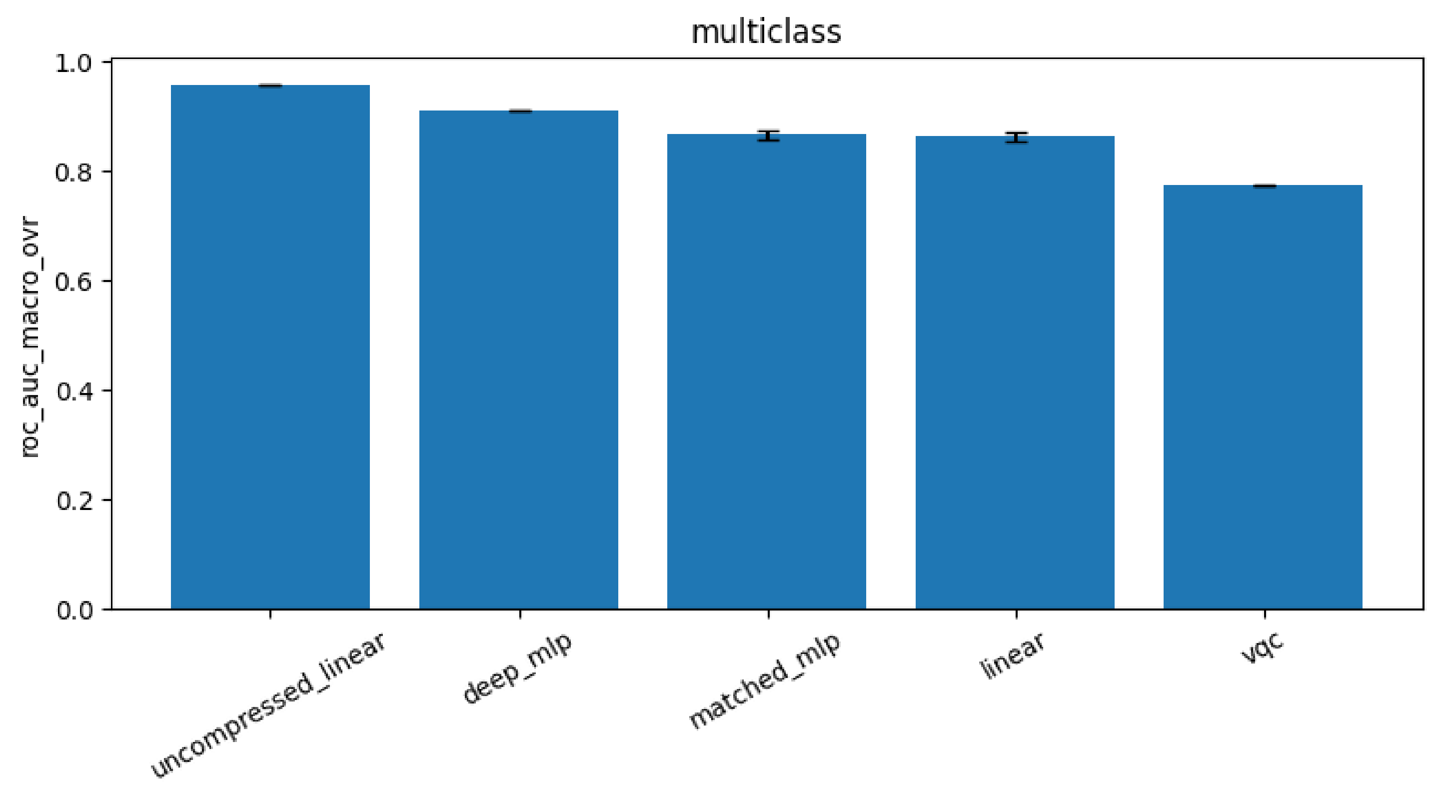


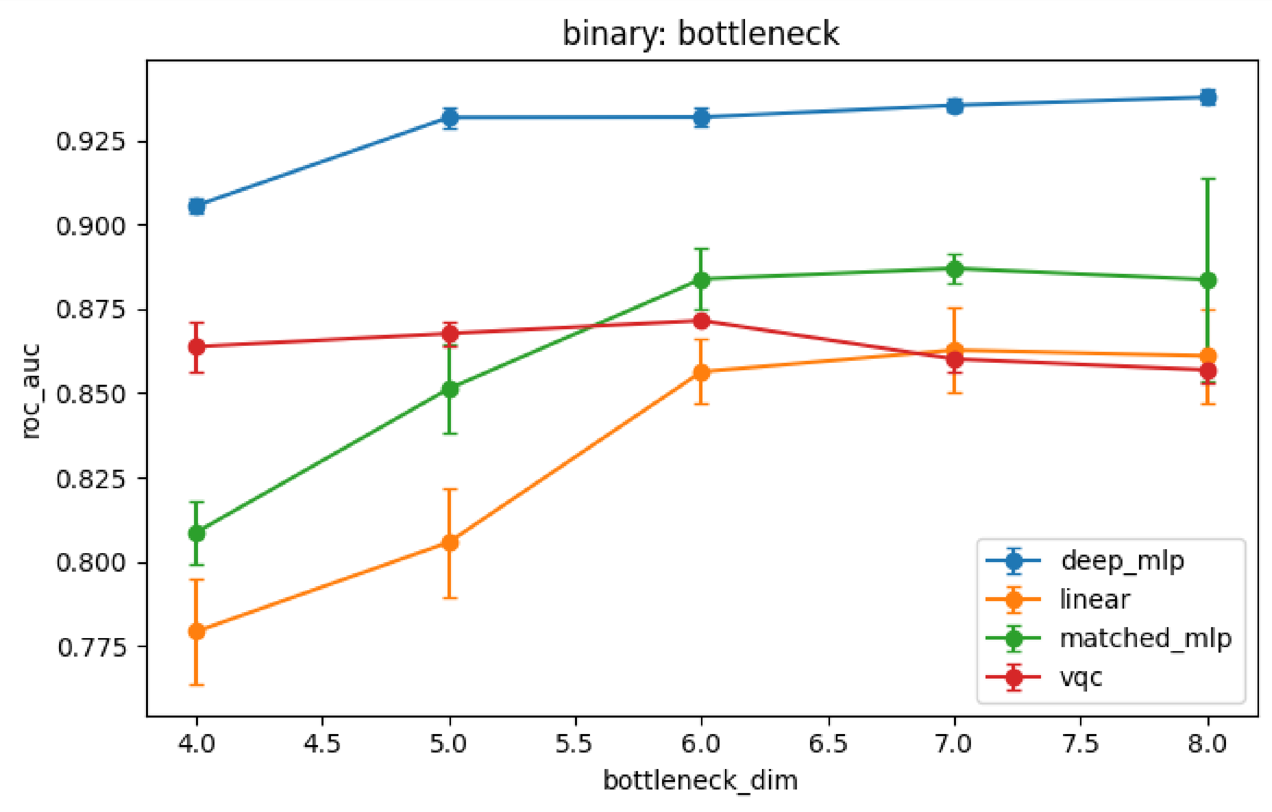
Fig 3.1 Multiclass Macro ROC-AUC

Results

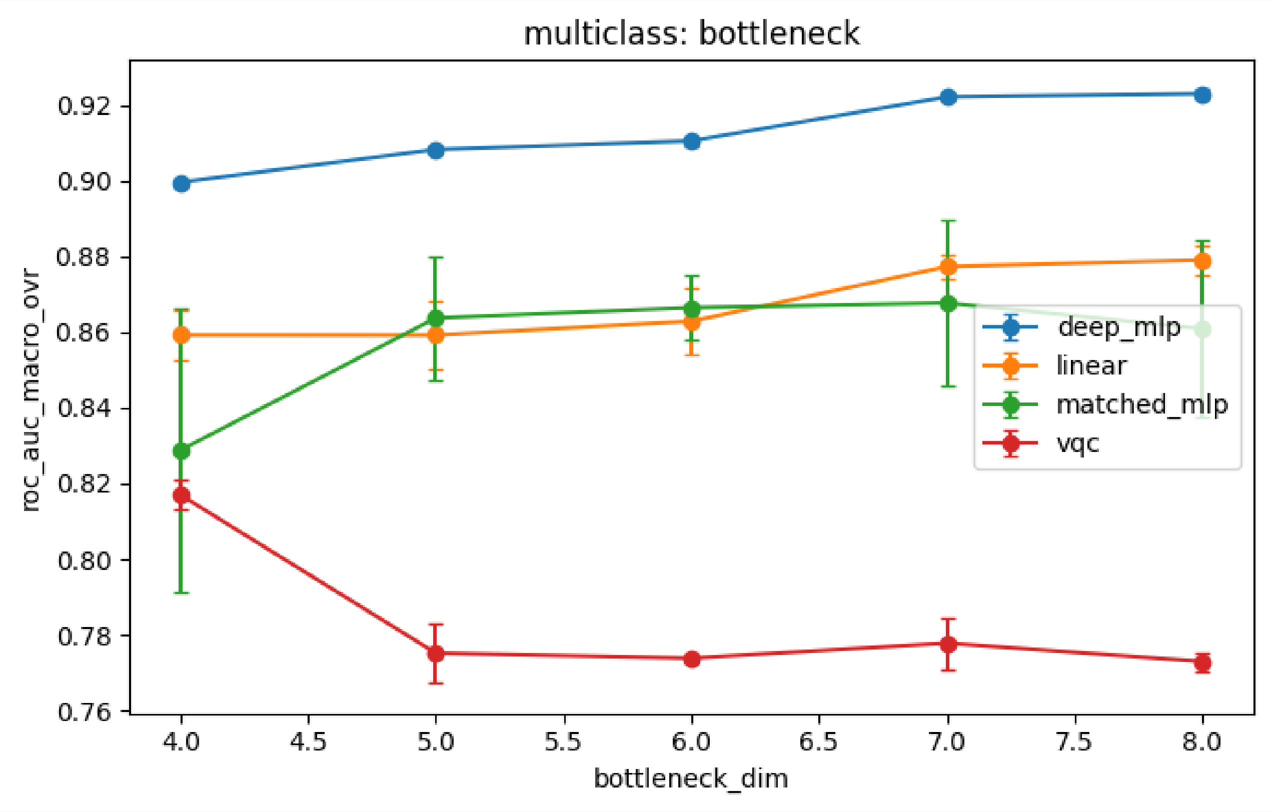
Head	Accuracy	ROC-AUC	ECE	Params
VQC	0.8038	0.8715	0.1019	36
Compressed linear	0.7788	0.8563	0.0609	14
Matched MLP	0.8030	0.8838	0.0719	38
Deep MLP	0.8562	0.9319	0.0222	2594
Uncompressed linear	0.9484	0.9899	0.0325	258

Fig 3.3 Accuracy

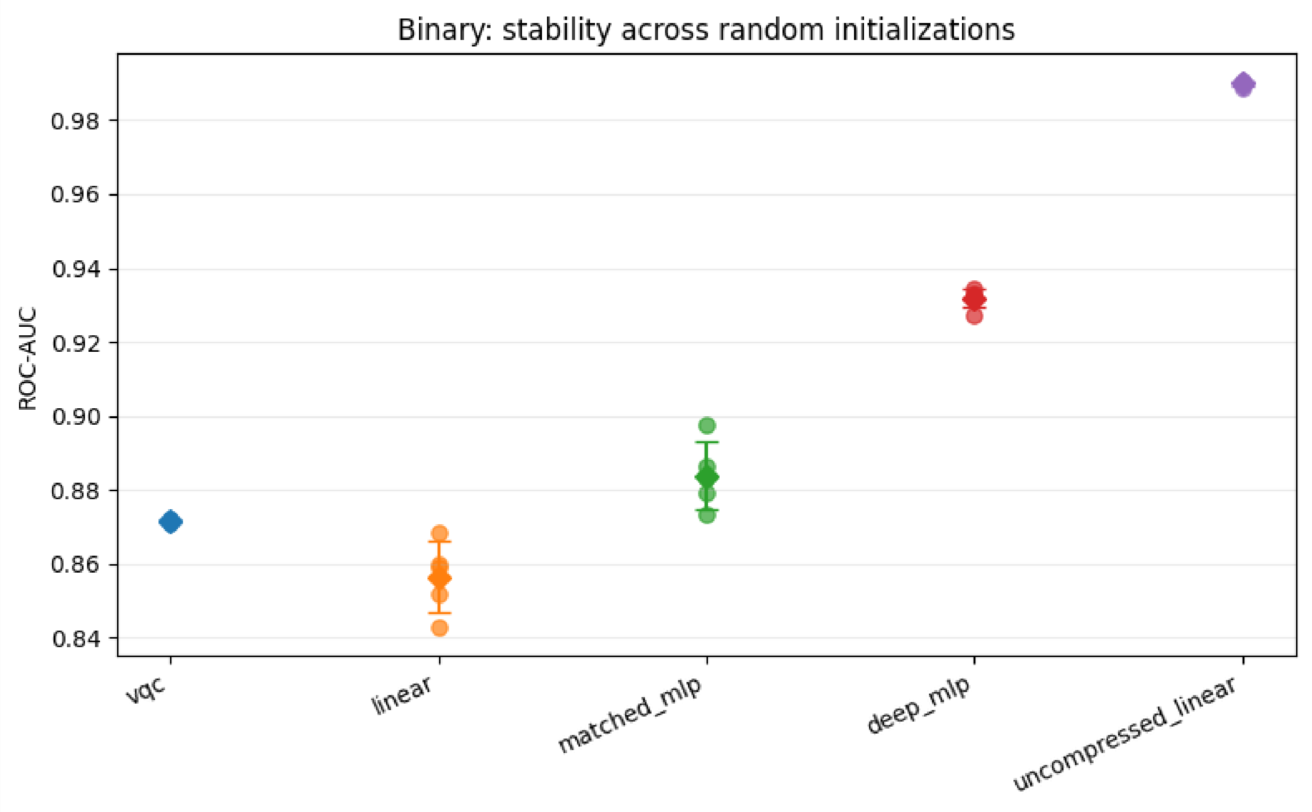
Controlled sweeps



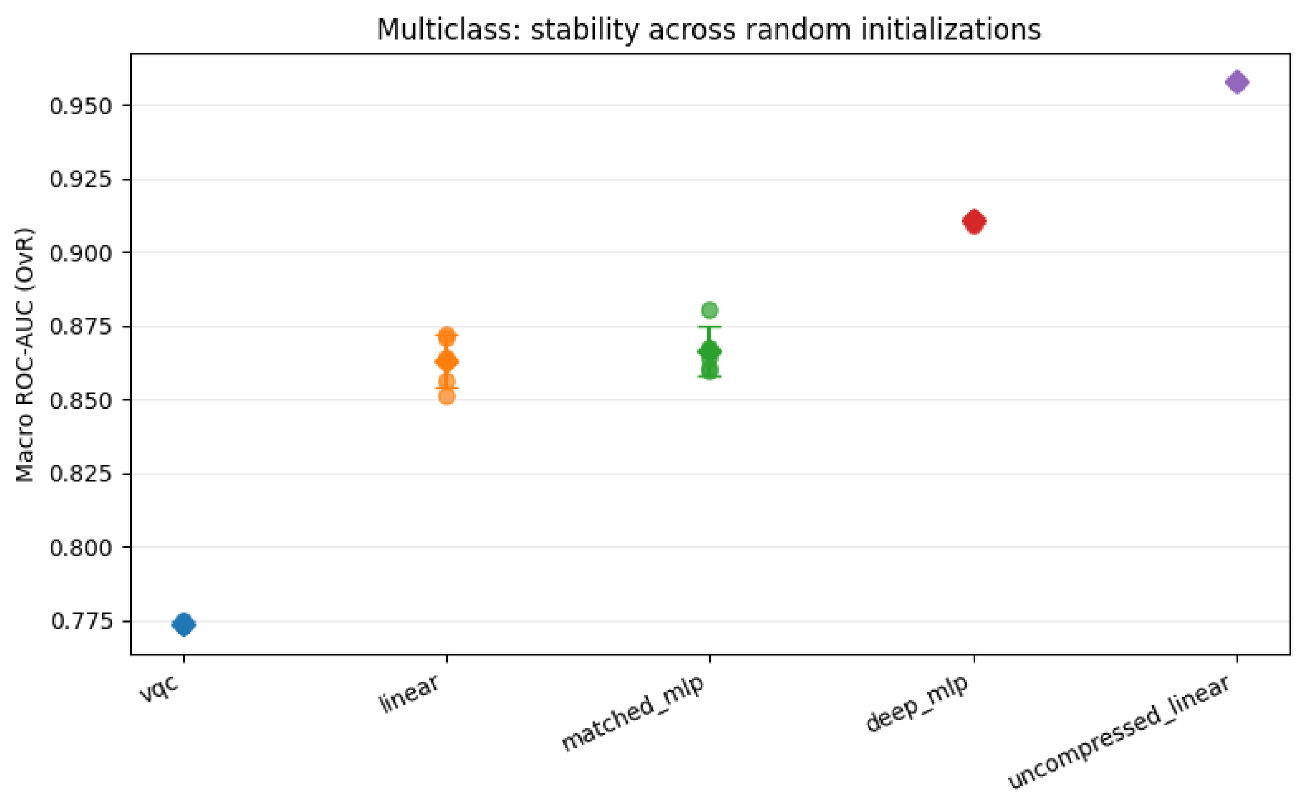
Performance



Performance



Performance



Takeaway

The aim is not to ask:

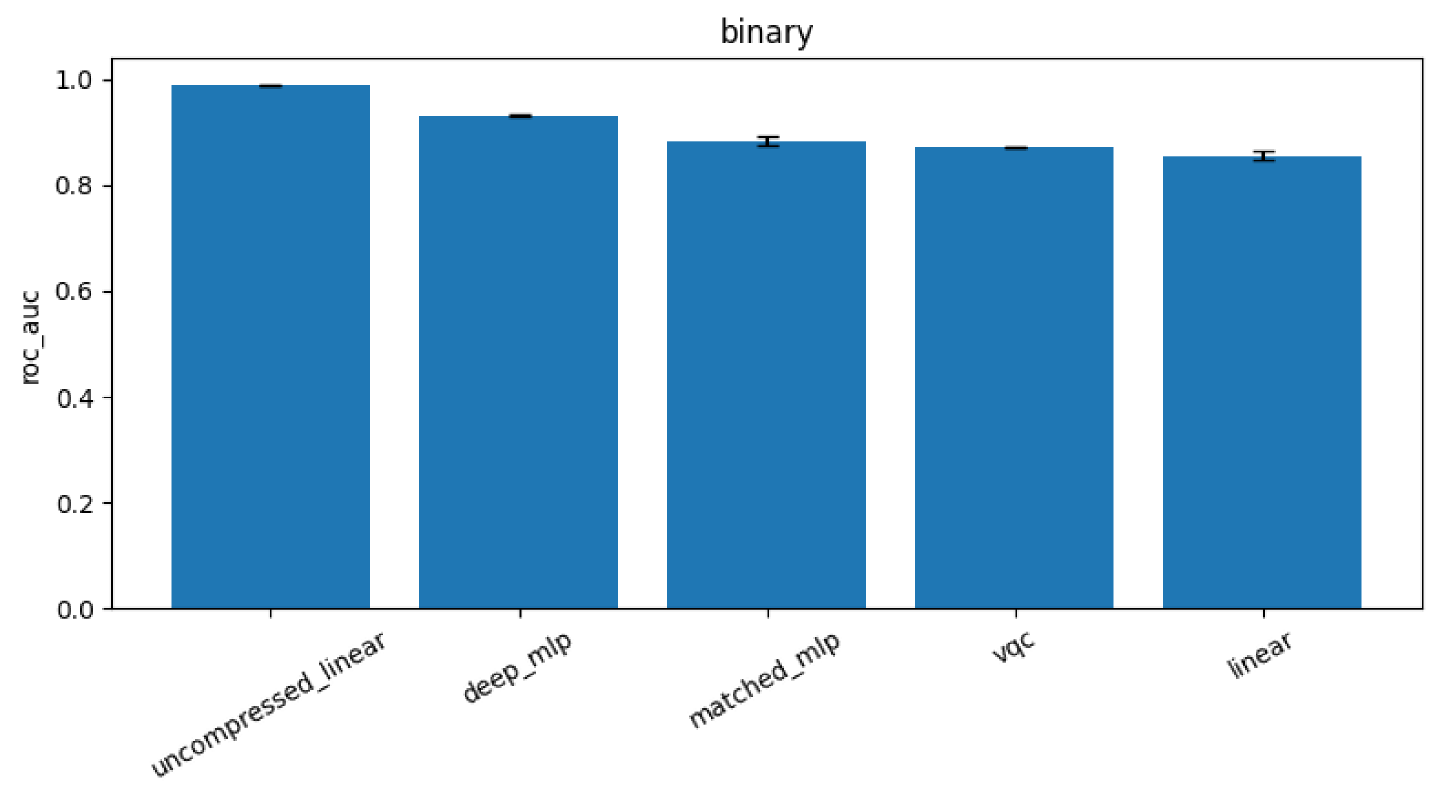
“*Did the quantum model win?*”

Instead, the study asks:

- Where is the VQC competitive with classical heads?
- Where do classical models retain a clear advantage?
- How much performance is limited by representation compression?
- Do more qubits, greater depth, or denser entanglement help?
- How sensitive are results to random initialization?
- What computational cost accompanies greater VQC complexity?

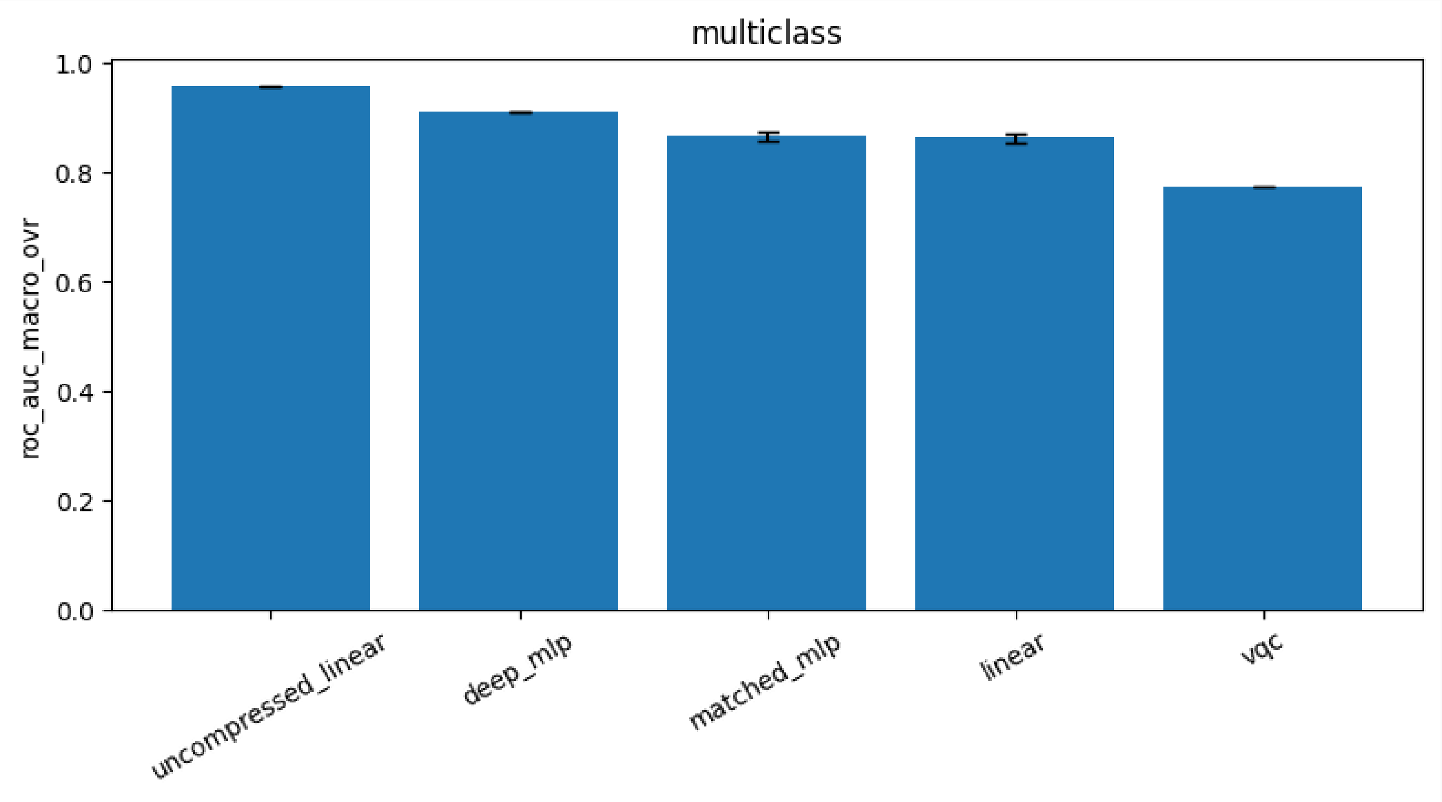
THANK YOU.

Backup

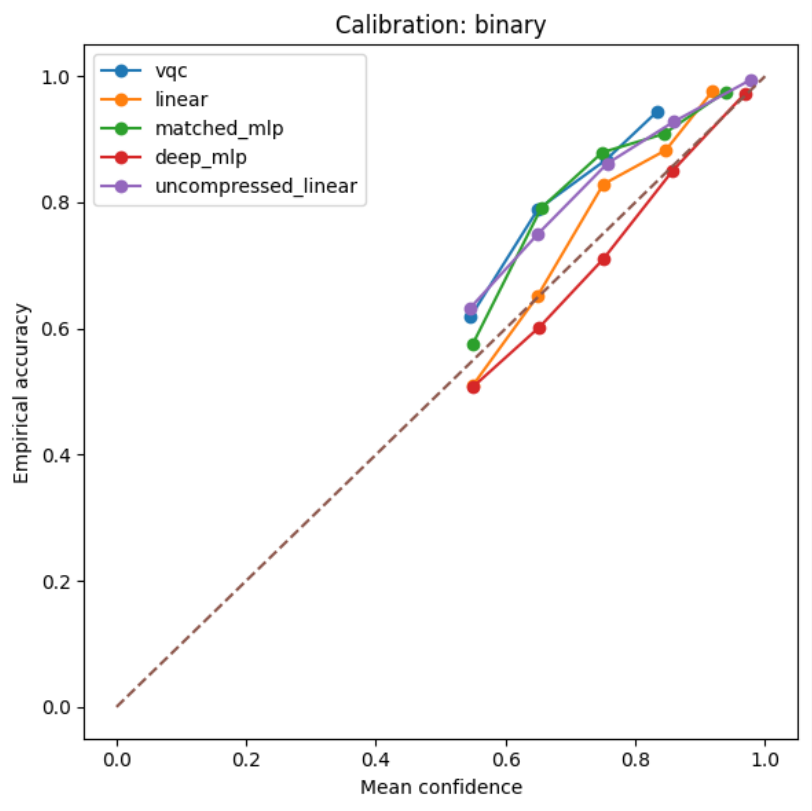


binary
roc_auc
1.0
0.8
0.6
0.4
0.2
0.0
uncompressed_linear
deep_mlp
matched_mlp
vqc
linear

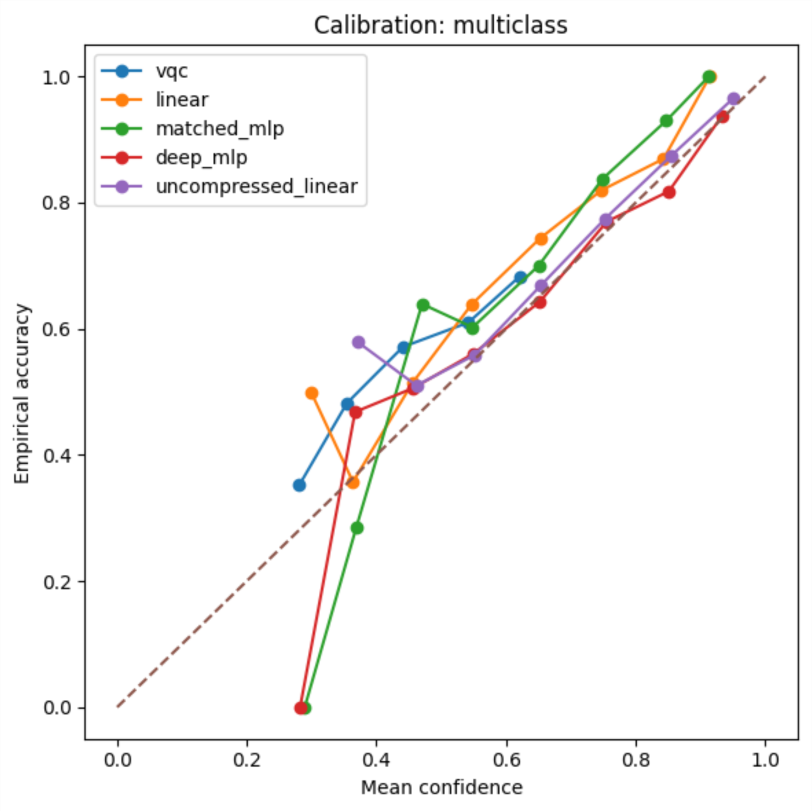
Backup


multiclass
roc_auc_macro_ovr
1.0
0.8
0.6
0.4
0.2
0.0
uncompressed_linear
deep_mlp
matched_mlp
linear
vqc

Backup

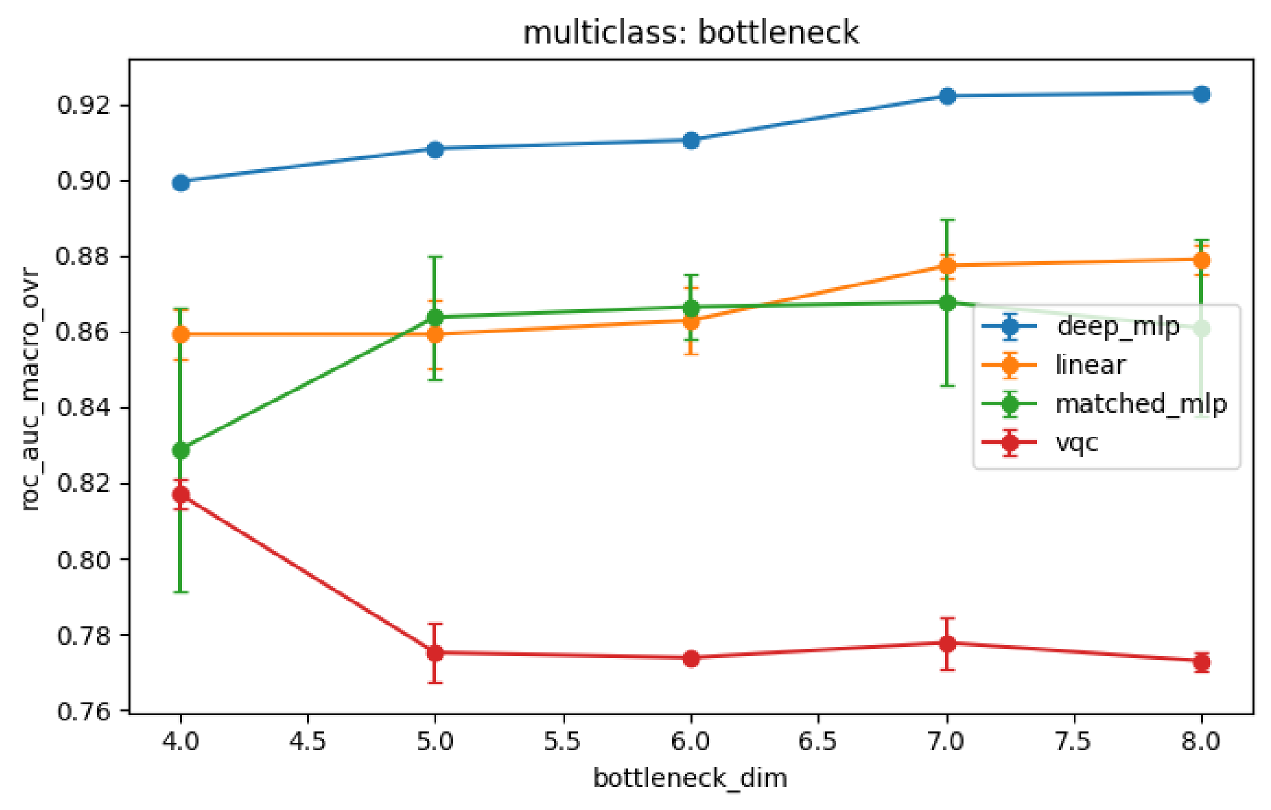


Calibration: binary
vqc
linear
matched_mlp
deep_mlp
uncompressed_linear
Empirical accuracy
Mean confidence
0.0
0.2
0.4
0.6
0.8
1.0



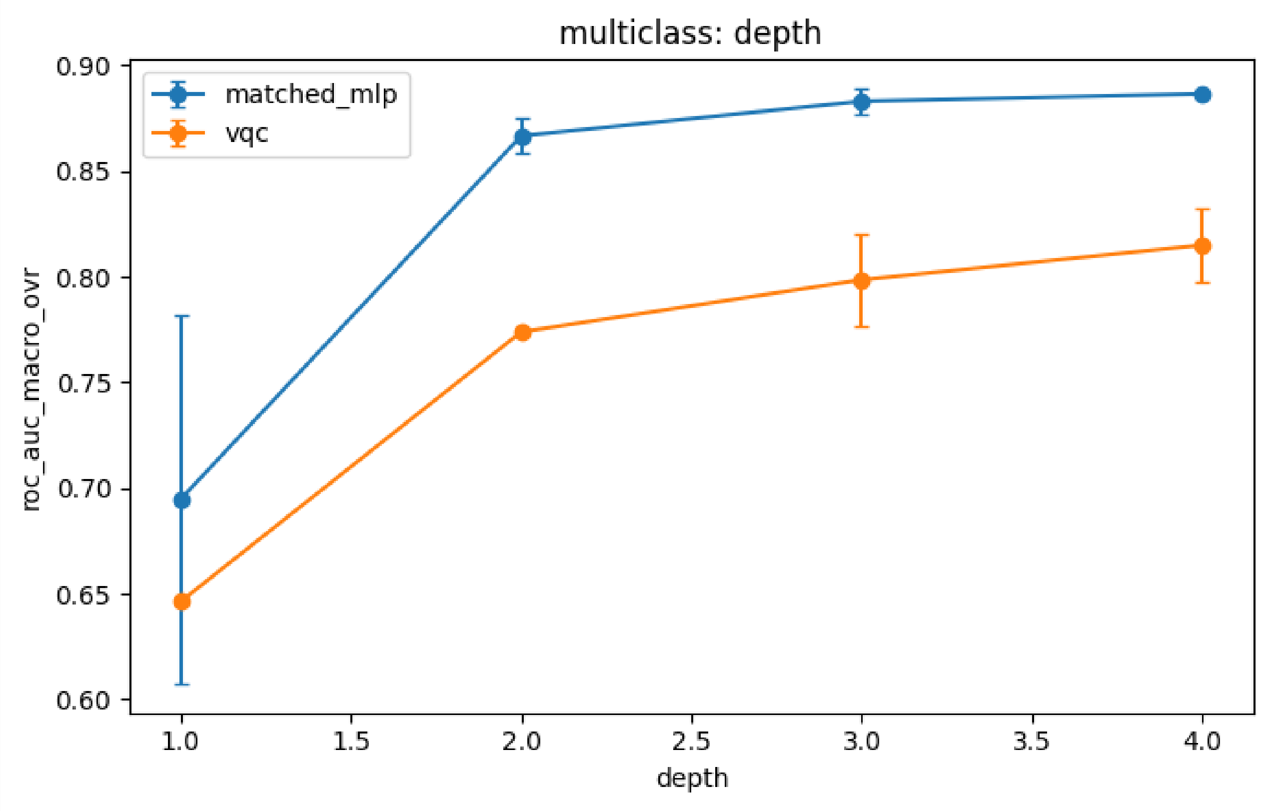
Calibration: multiclass
vqc
linear
matched_mlp
deep_mlp
uncompressed_linear
Empirical accuracy
Mean confidence
0.0
0.2
0.4
0.6
0.8
1.0

Backup



multiclass: bottleneck
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
roc_auc_macro_ovr
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
bottleneck_dim
deep_mlp
linear
matched_mlp
vqc

Backup



multiclass: depth
matched_mlp
vqc
0.90
0.85
0.80
0.75
0.70
0.65
0.60
roc_auc_macro_ovr
1.0
1.5
2.0
2.5
3.0
3.5
4.0
depth