Below is a detailed summary comparing three variations of your Pennylane-based hybrid quantum–classical models. Each version has been “working” (i.e. running without critical errors), but they differ in scope, architecture, and design choices. In particular, we compare:

1. **Binary Classification Model (Downsampled Input, 2 Classes)**
2. **10-Class Classification Model (Minimal Intrusive Changes from Binary Setup)**
3. **Expanded-Capacity 10-Class Model (Increased Quantum Capacity and Classical Preprocessing)**

**1. Binary Classification Model**

**Architecture Overview:**

* **Data Preprocessing:**
  + **Input Filtering & Downsampling:** The MNIST dataset was filtered to include only digits 0 and 1.
  + **Image Size:** Images were aggressively downsampled from 28×28 to 4×4 to reduce the number of pixels to be encoded.
* **Classical Feature Extraction:**
  + A relatively simple CNN (e.g., one or two convolution layers with pooling) is used to extract features from the 4×4 images.
  + The flattened features are then mapped via a Dense layer to a vector of length equal to the number of qubits (typically 4).
* **Quantum Circuit & Layer:**
  + **Embedding:** Uses AngleEmbedding to encode the compressed feature vector into qubit rotation angles.
  + **Variational Ansätze:** Uses a template (e.g. StronglyEntanglingLayers) with a modest circuit depth (for example, 3 variational layers).
  + **Output:** The quantum layer outputs expectation values for each qubit.
* **Final Classification:**
  + A Dense layer with 2 outputs (softmax) performs binary classification.

**Key Characteristics & Rationale:**

* **Simplicity:**
  + By reducing the task to a binary problem, the model focuses on a simpler decision boundary.
* **Resource Efficiency:**
  + Downsampling to 4×4 and using 4 qubits keeps the quantum part small and manageable.
* **Reported Performance:**
  + Research in hybrid models has shown that binary tasks can achieve very high accuracies (close to 99.7%) when the quantum circuit is well matched with the classical preprocessor.

**2. 10-Class Classification Model (Minimal Intrusive Changes)**

**Architecture Overview:**

* **Data Preprocessing:**
  + **Full Dataset:** Unlike the binary model, no filtering is performed—full MNIST (digits 0–9) is used.
  + **Image Size:** The original 28×28 resolution is maintained.
* **Classical Feature Extraction:**
  + **CNN Architecture:** A deeper classical CNN is used (more convolutional layers with padding and pooling) to extract richer features from the full-resolution images.
  + **Dense Mapping:** The CNN features are flattened and passed through a Dense layer to produce a vector whose length equals the number of qubits (e.g., 4).
* **Quantum Circuit & Layer:**
  + **Embedding & Variational Circuit:** The same QNode structure is used as in the binary case. It still uses AngleEmbedding and StronglyEntanglingLayers (with, say, 3 variational layers) over 4 qubits.
  + **Output:** The quantum layer outputs a feature vector of dimension 4.
* **Final Classification:**
  + A Dense layer with 10 outputs (softmax) is used to classify among the 10 digits.

**Key Characteristics & Rationale:**

* **Increased Complexity:**
  + Handling 10 classes is inherently more challenging. The classical CNN has been deepened to extract more informative features.
* **No Aggressive Downsampling:**
  + Using full-resolution images preserves detailed information—but it also increases the burden on the quantum circuit, which is still limited to 4 qubits.
* **Performance Trade-Off:**
  + Although the architecture is minimally modified from the binary prototype, the mismatch between full-resolution high-dimensional data and a small quantum circuit can limit accuracy relative to the binary case.
* **Simplicity in Changes:**
  + Compared to the binary model, the only modifications are that the filtering is removed, the input shape is restored to (28,28,1), and the final Dense layer is updated for 10 outputs. This is the least intrusive modification.

**3. Expanded-Capacity 10-Class Model**

**Conceptual Overview:**

*(Note: Although you requested “the least intrusive way” earlier, additional work has been done in research to scale hybrid models for multi-class tasks. This third model reflects those ideas.)*

* **Data Preprocessing:**
  + Uses the full MNIST dataset (28×28 images) without downsampling.
* **Enhanced Classical Feature Extraction:**
  + A more sophisticated CNN is employed—one that is deeper and possibly includes additional layers (or even residual connections) to distill a higher-dimensional feature vector from the images.
  + The output dense layer maps features to a higher-dimension vector (e.g., 8 or 16 dimensions) instead of matching a small number (like 4).
* **Quantum Circuit Adjustments:**
  + **Increased Qubit Count:** The number of qubits is raised from 4 to 8. This allows a larger Hilbert space, which is more suited for the richer (higher-dimensional) features coming from the deeper CNN.
  + **Variational Depth:** The quantum circuit may be altered (e.g., 2 layers or more) to balance between expressibility and trainability.
  + **Alternate Circuit Templates:** Research suggests that using specialized encoding (e.g., hybrid hierarchical encoding or customized amplitude/angle encoding) can benefit classification tasks. Although your current working code uses AngleEmbedding and StronglyEntanglingLayers, further experiments might incorporate a new ansatz inspired by the literature.
* **Final Classification:**
  + The output from the quantum layer (with a higher dimension) is passed into a Dense layer with 10 outputs (softmax).

**Key Characteristics & Rationale:**

* **Enhanced Quantum Capacity:**
  + Increasing the number of qubits and potentially the circuit depth allows the quantum component to encode more complex, higher-dimensional classical feature vectors.
* **Better Feature-Representation Matching:**
  + A richer classical feature extractor paired with a more expansive quantum circuit is intended to capture subtle differences required for accurate 10-class classification.
* **Trade-Offs:**
  + This model is more complex, both in terms of computational cost (more qubits, deeper network) and training difficulty. It requires careful tuning of hyperparameters (learning rate, circuit parameters, etc.) but is conceptually in line with research results showing that hybrid models can outperform classical CNNs if the quantum component is scaled properly.

**Comparative Summary**

* **Model 1 (Binary HQNN):**
  + *Task:* Binary classification (digits 0 and 1).
  + *Data:* Aggressively downsampled images (4×4).
  + *Architecture:* Simple classical CNN, 4 qubits in the quantum layer, 3 variational layers, and output of 2 classes.
  + *Performance:* Reports from literature show very high accuracy (~99.7%) under these simplified conditions.
* **Model 2 (10-Class HQNN with Minimal Changes):**
  + *Task:* Full 10-class classification.
  + *Data:* Full-resolution images (28×28).
  + *Architecture:* A classical CNN with moderate depth extracts features, which are mapped to a vector of length 4 (for 4 qubits) and passed through the same quantum circuit as before. The output dense layer has 10 units.
  + *Performance:* This model faces a mismatch between high-dimensional features and a limited (4-qubit) quantum circuit, making it less accurate than the binary model. It’s the least intrusive change—only altering the input shape and the final layer—so while it extends the task to 10 classes, performance might be lower compared with a model specifically designed for high-dimensional encoding.
* **Model 3 (Expanded-Capacity 10-Class HQNN):**
  + *Task:* 10-class classification.
  + *Data:* Full-resolution images (28×28).
  + *Architecture:* A deeper and more powerful classical CNN extracts richer features and maps them to a higher-dimensional vector (e.g., 8 or more dimensions). The quantum circuit is scaled up by increasing the number of qubits (e.g., to 8) and adjusting variational layers, better matching the classical feature vector’s complexity.
  + *Performance:* The increased quantum capacity is expected to better capture complex patterns and may push the hybrid model to higher accuracies, approaching or surpassing classical CNN performance if optimally tuned. However, increased complexity also leads to higher computational cost and demands more careful hyperparameter tuning.

**Conclusion**

* **Scope & Complexity:**
  + The binary model is simpler, has fewer parameters, and benefits from significant downsampling, resulting in high accuracy on a limited task.
  + The minimal 10-class model is the simplest extension but may suffer from limited quantum capacity relative to the complexity of the full MNIST task.
  + The expanded-capacity 10-class model attempts to bridge the gap by enhancing both classical feature extraction and quantum circuit capacity, at the cost of increased computational complexity and training challenges.

Each approach represents a trade-off between simplicity and capacity. The binary model is very efficient and accurate for two classes, while full 10-class classification requires scaling both the classical and quantum components. Research indicates that hybrid models can achieve extremely high accuracies when the system is balanced (i.e., when the quantum circuit is given enough capacity through additional qubits or deeper variational ansatz and paired with an effective classical feature extractor).

This detailed comparison should help guide further experiments based on your available resources and desired performance targets.