Journal of Quantum Technologies and Informatics Research
JQTAIR
KUANTUM TEKNOLOJİLERİ VE ENFORMATİK ARAŞTIRMALARI
JOURNAL OF QUANTUM TECHNOLOGIES AND INFORMATICS RESEARCH
Volume / Cilt: 3, Issue / Sayı: 1, 2025, e2702
E - ISSN: 3023-4735
URL: https://journals.gen.tr/index.php/jqtair/
DOI: https://doi.org/10.70447/ktve.2702

Received Date / Geliş Tarihi: 01/02/2025
Acccepted Date / Kabul Tarihi: 20/03/2025
Online Publication Date /
Online Yayın Tarihi: 07/04/2025

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

# Quantum-Enhanced Conformal Methods for Multi-Output Uncertainty: A Holistic Exploration and Experimental Analysis

**Davut Emre Tasar**

University of Navarra, Applied Engineering PhD, Pamplona, Spain, e-mail: dtasar@alumni.unav.es

## Abstract

*Quantum computing* introduces unique forms of randomness arising from measurement processes, gate noise, and hardware imperfections. Ensuring reliable *uncertainty quantification* in such quantum-driven or quantum-derived predictions is an emerging challenge. In classical machine learning, *conformal prediction* has proven to be a robust framework for distribution-free uncertainty calibration, often focusing on univariate or low-dimensional outputs [5]. Recent advances (e.g., [1–3]) have extended conformal methods to handle multi-output or multi-dimensional responses, addressing sophisticated tasks such as time-series, image classification sets, and quantum-generated probability distributions. However, bridging the gap between these powerful conformal frameworks and the high-dimensional, noise-prone distributions typical of quantum measurement scenarios remains largely open.

In this paper, we propose a unified approach to harness *quantum conformal methods* for multi-output distributions, with a particular emphasis on two experimental paradigms: (*i*) a standard 2-qubit circuit scenario producing a four-dimensional outcome distribution, and (*ii*) a multi-basis measurement setting that concatenates measurement probabilities in different bases (Z, X, Y) into a twelve-dimensional output space. By combining a multi-output regression model (e.g., random forests) with *distributional conformal prediction*, we validate coverage and interval-set sizes on both simulated quantum data and multi-basis measurement data. Our results confirm that classical conformal prediction can effectively provide coverage guarantees even when the target probabilities derive from inherently quantum processes. Such synergy opens the door to next-generation quantum-classical hybrid frameworks, providing both improved interpretability and rigorous coverage for quantum machine learning tasks. **All codes and full reproducible Colab notebooks** are made available at https://github.com/detasar/QECMMOU.

**Keywords:** Quantum Computing, Conformal Prediction, Multi-Output Regression, Distribution-Free Coverage, Multi-Basis Measurement, Quantum Machine Learning.

**Citation/Atıf:** TASAR, D.E. (2025). Quantum-Enhanced Conformal Methods for Multi-Output Uncertainty: A Holistic Exploration and Experimental Analysis. *Kuantum Teknolojileri ve Enformatik Araştırmaları*. 3(1): e2702, DOI: 10.70447/ktve.2702

**Corresponding Author/ Sorumlu Yazar:**
Davut Emre Tasar
E-mail: dtasar@alumni.unav.es

Bu çalışma, Creative Commons Atıf 4.0 Uluslararası Lisansı ile lisanslanmıştır.
This work is licensed under a Creative Commons Attribution 4.0 International License.

# 1 Introduction

**Motivation and Background.** Quantum computing leverages superposition and entanglement of qubits to (potentially) solve problems at scales intractable for classical machines. However, even as Noisy Intermediate-Scale Quantum (NISQ) devices progress [9], the inherent measurement randomness, gate errors, and hardware noise complicate the generation of stable output probabilities. Classical *conformal prediction* (CP) has emerged as a powerful, distribution-free framework that can quantify predictive uncertainty in a statistically rigorous way [4–6, 10]. Until recently, most conformal approaches addressed single-dimensional outputs or classification tasks [7]. Yet, the need to provide set-valued predictions or region-based coverage for multi-output data is rapidly growing [2, 8], especially in quantum contexts where the natural outputs (e.g., measurement distributions) are inherently multi-dimensional.

**Quantum-Specific Challenges.** Predicting the outcome distribution of a quantum circuit entails dealing with:

- *Stochasticity of measurement.* A 2-qubit circuit, for instance, yields a random distribution over $\{00, 01, 10, 11\}$. Each execution (shot) collapses the state, introducing inherent randomness.
- *Hardware noise and drifts.* Real quantum devices exhibit gate infidelities and drift over time, causing correlated noise across measurement shots [1, 9].
- *High-dimensional expansions.* If measurements are taken in multiple bases (e.g., X, Y, Z), the resulting multi-head distribution can reach dimension $4m$ for $m$ different bases in just a 2-qubit system.

Therefore, guaranteeing a coverage statement like "with 90% probability, the true measurement distribution lies within the predicted region" is non-trivial. Recent works [1] discuss *probabilistic conformal prediction (PCP)* for quantum models, but typically focus on single-basis or single-dimensional scenarios.

**Prior Art in Multi-Output Conformal.** Meanwhile, multi-output conformal methods in classical machine learning have flourished. For instance, ellipsoidal sets for multi-dimensional time series [2], multi-output regression intervals [3, 8], and adaptive or differentiable conformal solutions [2, 4] represent active directions. Additionally, works like [7] explore robust uncertainty quantification using Monte Carlo methods, while [8] provides insights into exact and approximate conformal inference for multi-output regression. Further contributions, such as [11], offer distribution-free predictive inference for regression, enhancing the theoretical foundation of multi-output methods. These techniques ensure finite-sample coverage under minimal assumptions—primarily *exchangeability* of calibration and test points. In quantum-like scenarios, exchangeability might hold when each circuit is drawn from the same distribution of gates or the same family of states, so the typical CP framework can be leveraged.

**Contributions and Paper Outline.** In this work, we propose a systematically integrated approach to:

1. Generate synthetic data from quantum circuits and multi-basis measurements. One scenario yields a 4-dim distribution from a 2-qubit measurement (computational basis), while the second scenario concatenates $Z$, $X$, and $Y$ measurement probabilities into a 12-dim vector.

2. Apply classical *multi-output regression* to map classical features (circuit-depth, gate counts, etc.) to these quantum-derived probability vectors.

3. Adopt a *distributional conformal prediction* approach, using a single scalar norm (e.g., $\ell_2$ or $\ell_\infty$) to form coverage sets in $\mathbb{R}^4$ or $\mathbb{R}^{12}$.

4. Experimentally evaluate how coverage changes with different miscoverage parameters $\alpha$, reporting coverage rates and "volume" (or set size) of these multi-dimensional intervals (hypercubes or hyperspheres).

We show that, under mild assumptions, conformal intervals indeed provide coverage close to $(1 - \alpha)$, even though the underlying data stems from quantum measurements. This synergy offers an important step toward quantum-classical hybrid pipelines where classical conformal methods supply robust uncertainty quantification for quantum outputs.

**Paper Organization.** We structure the paper as follows:

- **Section II (Materials and Methods)** details the quantum circuit generation, multi-basis measurement strategies, and the fundamentals of multi-output conformal intervals.
- **Section III (Experiments)** describes how we train random forest regressors on these multi-output distributions, and how we apply distributional conformal thresholds.
- **Section IV (Results and Discussions)** discusses coverage performance, the trade-off with set sizes, and potential limitations with quantum noise.
- **References and Appendices** provide literature context, expansions of conformal proofs, and the relevant code annotations for replicability.

By bridging modern quantum data generation with distribution-free uncertainty quantification, we illuminate a pathway to robust, "error-bounded" quantum machine learning—an essential milestone in advancing practical quantum computing.

# 2 Materials and Methods

In this section, we detail the *methodological* aspects of our study, which bridges quantum circuit data generation with multi-output conformal prediction. Specifically, we describe:

1. The *quantum data generation pipeline* for single-basis (2-qubit) as well as *multi-basis* measurements.

2. The *classical features* extracted from the quantum circuits.

3. Our *multi-output regression* strategy.

4. The *distributional conformal approach* for interval (or region) construction in up to 12 dimensions.

We also provide a conceptual flowchart in Fig. 1 to illustrate how these components interconnect.

## 2.1 Overview of the Pipeline

Figure 1 summarizes the overarching pipeline, from quantum circuit generation to final conformal sets. The pipeline is divided into four major phases:

1. **Circuit and Measurement Setup:** We design a random 2-qubit quantum circuit, optionally measuring it in one or more bases (e.g., $\{Z, X, Y\}$).
2. **Data Extraction:** We gather classical features (gate counts, depth, etc.) in an $\boldsymbol{X}$ matrix, along with measured distribution vectors (probabilities) in a $\boldsymbol{Y}$ matrix.
3. **Model Training:** A multi-output regression model is trained to approximate $f : \boldsymbol{X} \to \boldsymbol{Y}$.
4. **Conformal Inference:** Using a calibration subset, we derive coverage thresholds (radii) for distributional conformal sets in $\mathbb{R}^4$ or $\mathbb{R}^{12}$. We then test coverage on new circuits.

## 2.2 Quantum Circuit Generation

**2-Qubit Architecture.** We focus on 2-qubit circuits due to their simplicity and ability to demonstrate non-trivial entanglement. Each qubit is initialized to $|0\rangle$. The circuit depth $D$ (ranging from 1 to 8 in some experiments) dictates the number of gate layers. A depth-1 circuit could apply one or two gates (including possible controlled gates), whereas a depth-8 circuit can have up to 16 or more operations, depending on the random draws.

**Random Gate Selection.** Following [1] and others, we employ a pseudorandom scheme (e.g., Qiskit's `random_circuit()`) that draws from common gates (`H, X, Y, Z, RX, RZ, CX, ...`). This ensures a diverse distribution of unitary transformations. At the end of each circuit, we apply a measurement operation:

- *Single-Basis (Z) scenario:* We measure in the computational (Z) basis, obtaining probabilities for $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.
- *Multi-Basis scenario:* We replicate the circuit or apply basis transformations to measure in $X$, $Y$, or $Z$ bases. For instance, measuring in the $X$ basis typically involves a Hadamard on each qubit prior to a $Z$-basis readout. Similarly, measuring in the $Y$ basis can be done via $S^\dagger$ then $H$ [9].

Each measurement returns a bitstring from $\{00, 01, 10, 11\}$ for 2 qubits, repeated over $N_{\text{shots}}$ runs to estimate probabilities.

## 2.3 Feature Extraction

Every circuit is analyzed to yield a vector of classical features, $\mathbf{x} \in \mathbb{R}^m$, capturing:

- **Depth (integer):** $D \in \{1, 2, \ldots, 8\}$.
- **Total gates:** Summation of all single- and two-qubit gates used.
- **Gate counts:** For each gate type (H, X, Y, Z, CX, ...), we store how many times it appears.

In our simplest multi-basis experiment, we used $m = 2$ features: $\{D, \text{total_ops}\}$ for brevity. One can easily extend it to $m = 17$ by enumerating all gate-type frequencies.

## 2.4 Measurement Vectors

**Single-Basis Setup (4D).** Measuring only in the Z-basis yields four probabilities $(p_{00}, p_{01}, p_{10}, p_{11})$, which sum to 1. We store them as a 4D vector $\mathbf{y} \in [0,1]^4$. In practice, the model may predict a 4D vector that does not sum to 1; we do not impose an explicit simplex constraint in our regressor, but it remains a mild approximation.

**Multi-Basis Setup (12D).** When measuring in $Z$, $X$, and $Y$ bases, each basis yields four probabilities $\mathbf{z}, \mathbf{x}, \mathbf{y} \in [0,1]^4$ (not to be confused with the notation for circuit features or for $\mathbf{y}$ as ground truth). We concatenate these results:

$$\mathbf{Y}_{\text{multi}} = (p^Z_{00}, p^Z_{01}, p^Z_{10}, p^Z_{11},\ p^X_{00}, p^X_{01}, p^X_{10}, p^X_{11},\ p^Y_{00}, p^Y_{01}, p^Y_{10}, p^Y_{11}).$$

Hence, $\mathbf{y} \in \mathbb{R}^{12}$. This approach yields a richer "multi-head" measurement vector.

## 2.5 Multi-Output Regression Model

To predict $\mathbf{y}$ from the classical feature vector $\mathbf{x}$, we use a multi-output random forest regressor [5] by default. Each tree in the ensemble splits on $\mathbf{x}$, and outputs a (4- or 12)-dim leaf average, aggregated across the forest. This classical approach allows straightforward handling of moderately large feature vectors (e.g., 17 gates) and multi-dimensional outputs. Our primary interest is *not* to achieve minimal MSE but to demonstrate how conformal intervals adapt to model accuracy. One may replace random forests with neural networks, gradient boosting, or quantum-hybrid regressors [3,4] with minimal pipeline changes.

## 2.6 Distributional Conformal Inference

Let $f(\mathbf{x})$ be the trained regressor. For a calibration set $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, define a residual metric:

$$r_i = \|\mathbf{y}_i - f(\mathbf{x}_i)\|_\infty \quad \text{or} \quad r_i = \|\mathbf{y}_i - f(\mathbf{x}_i)\|_2, \tag{1}$$

depending on whether an $\ell_\infty$ or $\ell_2$ norm is desired. We sort these $r_i$ in ascending order to get $r_{(1)} \le r_{(2)} \le \cdots \le r_{(n)}$. For a user-chosen $\alpha \in (0,1)$, let $k = \lceil (1-\alpha)\,(n+1) \rceil$. Then

$$\tau_\alpha = r_{(k)}.$$

Given a new test point $\mathbf{x}_{\text{new}}$, the conformal set is

$$\mathcal{C}_\alpha(\mathbf{x}_{\text{new}}) = \big\{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - f(\mathbf{x}_{\text{new}})\| \le \tau_\alpha\big\}, \tag{2}$$

where $d$ is 4 in single-basis or 12 in the triple-basis scenario. This set is typically a hypercube (for $\ell_\infty$) or hypersphere (for $\ell_2$). By exchangeability arguments, the true $\mathbf{y}_{\text{new}}$ should lie in $\mathcal{C}_\alpha(\mathbf{x}_{\text{new}})$ with probability at least $(1-\alpha)$ in finite samples [6].

**Coverage and Set Size.** We measure coverage on the test set by checking $\mathbf{y}_j^{(\text{test})} \in \mathcal{C}_\alpha(\mathbf{x}_j^{(\text{test})})$ for each sample $j$. The fraction of "inside" events approximates the coverage. Meanwhile, the "size" can be measured by volume (e.g., $(2\tau_\alpha)^d$) or sum-size (like $d \cdot 2\tau_\alpha$ in $\ell_\infty$ norms). This trade-off between coverage $(1-\alpha)$ and set-size is central to conformal methods.

**Exchangeability in Quantum Data.** While quantum hardware might exhibit correlated noise across runs, our synthetic approach typically re-seeds each circuit, so each example $\mathbf{x}_i$ is exchangeably drawn from a circuit distribution. If calibration and test sets are similarly drawn, the fundamental assumption for conformal validity holds. Of course, real quantum hardware drifts remain an open challenge, and advanced PCP solutions [1] might handle time-varying noise.

## 2.7 Implementation Details and Code Structure

Our public code is separated into two core notebooks:

- *Data Generation Notebook*: Implements the procedures in Sec. 2.2–2.4 for single- or multi-basis measurements. Stores $(\mathbf{X}, \mathbf{Y})$ arrays in `.npz` or `.pkl` files.
- *Model + Conformal Notebook*: Loads the data, splits into train/cal/test, trains a multi-output regressor (random forest), computes residuals, and evaluates coverage for various $\alpha$.

A typical run might produce coverage near $(1 - \alpha)$ with 4D data, and similar or slightly larger coverage sets with 12D multi-basis data. For completeness, *Listing 1* (pseudo-code) outlines our Pythonic pipeline:

**Listing 1.** Pseudo-code for quantum data generation and distributional conformal.

```
# (A) Data Generation
Define num_samples, min_depth, max_depth, shots.
X_list, Y_list = [], []

for i in range(num_samples):
qc = random_circuit(num_qubits=2, depth=rand_in_[min_depth,max_depth])
x_features = extract_gate_counts(qc)
measure_Z = get_distribution(qc, basis='Z')
measure_X = get_distribution(qc, basis='X')
measure_Y = get_distribution(qc, basis='Y')
y_vector = concat(measure_Z, measure_X, measure_Y)
X_list.append(x_features), Y_list.append(y_vector)

X_data = np.array(X_list); Y_data = np.array(Y_list)
save(X_data, Y_data, filename=...)

# (B) Conformal Pipeline
X_train, X_cal, X_test, Y_train, Y_cal, Y_test = splits(...)
model = RandomForestRegressor(...).fit(X_train, Y_train)
cal_pred = model.predict(X_cal)
resid_list = [ norm(Y_cal[i] - cal_pred[i]) for i in range(len(cal_cal)) ]
sort_resid = sorted(resid_list)

for alpha in [0.05, 0.10, 0.20, 0.30, 0.50]:
idx = ceil((1-alpha)*(len(cal_cal)+1)) - 1
tau = sort_resid[idx]
# Evaluate coverage on test
test_pred = model.predict(X_test)
coverage = mean( [norm(Y_test[j]-test_pred[j]) <= tau] )
print("alpha=", alpha, " coverage=", coverage, " tau=", tau)
```

The above references `measure_Z`, `measure_X`, `measure_Y` as distinct measurement routines. In a real experiment, we either re-initialize the circuit or suitably rotate qubits prior to measuring them in each basis.
This completes the **Materials and Methods** section. Next, we detail the specific *experimental setups*, hyper-parameters, and further *results* that validate or highlight the coverage properties of this approach.

# 3 Experimental Setup

In this section, we detail the experimental configurations designed to showcase the *quantum-to-classical* data generation process and the subsequent application of *distributional conformal prediction*. Our experiments aim to demonstrate two primary scenarios:

- **Single-Basis (4D) Approach:** Measuring a 2-qubit circuit only in the computational ($Z$) basis. Each circuit instance yields a single 4D probability vector.
- **Multi-Basis (12D) Approach:** Measuring the same circuit in three distinct bases ($Z$, $X$, and $Y$), concatenating three 4D distributions for each circuit, thus producing 12-dimensional outputs.

We first describe the data sources, including toy and large-scale datasets. Next, we outline how *train–cal–test* splits are performed, followed by the key experimental hyper-parameters. Finally, we summarize the procedure for evaluating coverage in both 4D and 12D scenarios.

## 3.1 Datasets

**Toy Dataset (Hundreds of Samples).** We prepared a smaller dataset—on the order of 200–500 circuits—to allow quick debugging, real-time visualizations, and faster iteration on code. In this dataset:

- The circuit depth $D$ was randomly sampled from $\{1, 2, 3, 4\}$.
- The random gate set typically included $\{H, X, Y, Z, \mathrm{CX}, \ldots\}$ as described in Sec. 2.2.
- Number of shots was set to 512 or 1024 to maintain moderate precision in measuring probabilities.
- For multi-basis generation, we measured the same circuit across $Z, X, Y$ by applying the relevant transformations (Hadamard, $S^\dagger$, ...).

Due to the lower sample count, these toy datasets are not intended to reflect large-scale performance but rather to illustrate the pipeline's mechanics for a random circuit example).

**Large-Scale Dataset (Tens of Thousands of Samples).** A second dataset of size 20000 or 50000 circuits was generated to examine more robust coverage properties. This larger-scale set:

- Used circuit depths up to 8, potentially reaching up to 16 or more gates.

- Was subject to a duplicate-dropping step (Sec. 2.3), ensuring we do not store the same gate composition multiple times.
- Provided enough calibration/test data to yield stable coverage estimates even for small $\alpha$ (like $\alpha = 0.05$).

## 3.2 Train–Cal–Test Splits

After data generation, each dataset $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N}$ is randomly partitioned into three disjoint subsets:

- **Training Set:** Roughly 70% of the data, used for fitting the multi-output regressor.
- **Calibration Set:** Around 15% of the data, used to compute the conformal residual thresholds (Sec. 2.6).
- **Test Set:** The remaining 15%, reserved for final evaluation of coverage and set size.

Because we rely on exchangeability for theoretical validity, each sample is drawn by an i.i.d. seed for random circuit generation. In the multi-basis scenario, each sample $\mathbf{y}_i \in \mathbb{R}^{12}$ arises from measuring that random circuit in three bases. If hardware drift or correlated shot noise existed, partial violation of exchangeability might occur [1], but for these simulator-based tests, the assumption holds neatly.

## 3.3 Hyperparameters and Implementation Details

**Quantum Circuit Parameters.** We typically fix the number of qubits to $n = 2$. For each sample:

- *Depth D*: Uniformly drawn in $\{1, \ldots, 8\}$ or $\{1, \ldots, 4\}$ depending on the experiment size.
- *Shots* # = 1024: This is the default for most runs; lower shots (256) or higher (2048) can be used to explore noise/variance trade-offs.

**Classical Feature Extraction.** We tested two schemes:

1. *Minimal Features* $(D, \text{total_ops})$, yielding a 2D $\mathbf{x}$.
2. *Full Gate Counts* $(D, \text{total_ops}, \text{count_H}, \ldots)$, which can be 17D or higher.

In all cases, duplicates are removed by a `pandas.drop_duplicates` based on feature columns.

**Random Forest Regressor.** We use #trees = 50 or #trees = 100 with default MAX_DEPTH as None. We set RANDOM_STATE = 42 for reproducibility. No advanced hyperparameter tuning is performed, as the aim is to illustrate coverage, not to minimize MSE.

**Residual Metric.** Unless stated otherwise, we adopt the $\ell_2$ (Euclidean) norm for multi-basis 12D experiments:

$$r_i \;=\; \big\|\mathbf{y}_i - \hat{\mathbf{y}}_i\big\|_2 \;=\; \sqrt{\sum_{j=1}^{d} \big(y_{i,j} - \hat{y}_{i,j}\big)^2},$$

where $d \in \{4, 12\}$ depending on single- or multi-basis outputs. For single-basis 4D experiments, we sometimes tested $\ell_\infty$ to see differences in coverage set shape.

## 3.4 Experiment Configurations

### 3.4.1 Single-Basis 4D Distributions

Here, each sample is a single circuit measured in the Z-basis. The output $\mathbf{y}_i \in \mathbb{R}^4$. After training the model, a calibration set of size $\approx 0.15N$ is used to gather residuals. Then:

1. We pick $\alpha$ from $\{0.05, 0.1, 0.2, 0.3, 0.5\}$.
2. Compute $\tau_\alpha$ using the formula in eq. (1) plus the sorted residual approach.
3. Evaluate coverage on the test set.

The coverage set $\mathcal{C}_\alpha(\mathbf{x}) \subset \mathbb{R}^4$ forms a 4D ball or hypercube around the predicted point, depending on the chosen norm.

### 3.4.2 Multi-Basis 12D Distributions

For the triple-basis approach, we run each circuit and measure $\{Z, X, Y\}$ bases, generating a 12D output vector $\mathbf{y}$. This is repeated up to $N = 20000$ to have robust coverage estimates. The same train–cal–test procedure applies, except we are now in $\mathbb{R}^{12}$. While coverage remains near $(1-\alpha)$, the conformal sets necessarily become higher dimensional, often requiring larger radii to encapsulate the same fraction of test points.

**Motivation for Multi-Basis.** Multi-basis distributions can reveal diverse quantum properties. For instance, a circuit that yields near-certain $|00\rangle$ in $Z$-basis might yield fairly uniform outcomes in $X$- or $Y$-basis. By combining them, we present a more holistic picture of how the circuit transforms states under different measurement contexts [9]. This can be especially relevant if the user wants *all* basis outcomes to be predicted with guaranteed coverage.

## 3.5 Performance Metrics

**Coverage and Set Size.** As in standard conformal literature [4, 6], for each chosen $\alpha$:

- **Empirical Coverage**: The fraction of test points $\mathbf{y}_j$ that satisfy $\|\mathbf{y}_j - \hat{\mathbf{y}}_j\| \le \tau_\alpha$.
- **Radius** $\tau_\alpha$: The scalar threshold derived from calibration.
- **Sum-Size / Volume** (optional): For an $\ell_\infty$ ball in dimension $d$, the sum-size is $2d\,\tau_\alpha$, while the volume is $(2\tau_\alpha)^d$. For $\ell_2$-balls, the $d$-dim volume formula is $\pi^{d/2}\,\tau_\alpha^d / \Gamma(\frac{d}{2}+1)$.

We typically display coverage vs. $\alpha$ and also examine how large $\tau_\alpha$ grows for small $\alpha$.

**Mean Squared Error.** Although the random forest regressor is not necessarily optimized for minimal MSE, we still report it to gauge how well the classical model approximates quantum measurement distributions. MSE can be computed per dimension or as a uniform average across $d$ outputs. If MSE is large, we expect bigger $\tau$ to maintain coverage.

### 3.6 Summary of the Experimental Setup

Overall, the experiments described in this section are structured as follows:

1. **Generate Data:** either single- or multi-basis, with a user-defined number of samples. Remove duplicates in features.
2. **Split:** $\approx 70\%$ train, $\approx 15\%$ calibration, $\approx 15\%$ test.
3. **Train Regressor:** multi-output random forest with moderate hyperparameters.
4. **Calibrate Residuals:** Sort $\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|$ on the calibration set to get $\tau_\alpha$ at each $\alpha$.
5. **Evaluate:** coverage, radius, volume, or sum-size on the test set.

As we will see in the next section, *Results and Discussion*, this setup reliably demonstrates the characteristic trade-off between $\alpha$ and coverage for quantum circuit data, and reveals how multi-basis output predictions can effectively be given coverage guarantees in up to 12 dimensions.

## 4 Results and Discussion

In this section, we systematically present the outcomes obtained from applying both *single-basis* (4D) and *multi-basis* (12D) data generation to the conformal prediction pipeline. Our main focus lies on evaluating how coverage and set sizes (i.e., the resulting multi-dimensional "uncertainty" regions) behave under varying miscoverage levels $\alpha$. We also compare model accuracy in terms of mean squared error (MSE), emphasizing how a quantum data source can challenge classical regressors if the resulting circuit distributions exhibit high complexity.

### 4.1 Single-Basis (4D) Coverage

**Coverage *vs.* $\alpha$.** When each 2-qubit circuit is measured solely in the computational ($Z$) basis, the outputs are four-dimensional vectors $(p_{00}, p_{01}, p_{10}, p_{11})$. Across various values of $\alpha \in \{0.05, 0.10, 0.20, 0.30, 0.50\}$, we observe a consistent pattern in which the *empirical coverage* on the test set is at or above the desired nominal coverage $(1-\alpha)$. For instance, if $\alpha = 0.10$, coverage near 90% is typically reached, confirming that our distributional conformal approach meets its theoretical design objective, in line with the finite-sample guarantees established in existing conformal literature [1, 6].

**Set Size.** In the 4D scenario, adopting either an $\ell_2$ or $\ell_\infty$ norm for residual computation entails different geometric shapes of the conformal set:

- With $\ell_\infty$, the conformal set is a 4D hypercube of edge length $2\,\tau$.
- With $\ell_2$, it is a 4D hypersphere (radius $\tau$).

Quantitatively, smaller $\alpha$ values ($\alpha \leq 0.10$) yield fairly large sets, reflecting the model's need to enclose a higher fraction of points. Meanwhile, for moderate $\alpha \geq 0.30$, the radius $\tau$ declines, sometimes drastically, but coverage likewise dips toward 70% or lower.

**Random-Forest MSE.** We typically observe an overall MSE in the 0.05–0.08 range, depending on circuit depth distribution and shot noise. Certain bitstrings (e.g., $p_{01}$ or $p_{11}$) may have slightly higher dimension-wise MSE, since random circuits do not always produce uniform or easily predictable patterns. Nonetheless, the moderate MSE is sufficient for the conformal procedure to yield coverage near $(1 - \alpha)$.

## 4.2 Multi-Basis (12D) Coverage

**Why 12D?** In the multi-basis case, each circuit is measured in $Z$, $X$, and $Y$ bases, concatenating three separate (4)-element probability distributions into a 12-dimensional vector. This approach exposes the regressor to a broader characterization of the quantum state, but also increases the complexity of the learning task [3, 4].

**Empirical Coverage Trends.** Despite the jump to a higher-dimensional output space, our results consistently show that distributional conformal sets still achieve near-nominal coverage: for $\alpha = 0.05$, coverage often lands around 95%–97%; for $\alpha = 0.10$, coverage remains around 90%. However, this comes at the cost of a larger residual radius $\tau$. In 12D, the single-radius ball (under $\ell_2$-norm) may require $\tau \approx 0.8$ or higher to envelop enough calibration samples. A typical coverage table reads something like:

$$(\alpha = 0.05 \Rightarrow \text{coverage} = 0.95), \quad (\alpha = 0.10 \Rightarrow \text{coverage} = 0.91), \quad (\alpha = 0.30 \Rightarrow \text{coverage} = 0.70).$$

One can trace these values to the complexity of simultaneously matching circuit distributions for three measurement bases, each of which can vary widely depending on circuit entanglement and global phases [1, 2].

**Comparisons of Set Size.** Because $\ell_2$-balls in 12D can exhibit dramatically larger volume than 4D balls for the same radius, we often monitor coverage *and* radius rather than volume. Indeed, even a moderate radius can produce extremely large volumes in 12D. As $\alpha$ increases, the needed radius declines (e.g., from $\tau \approx 1.05$ at $\alpha = 0.05$ down to $\tau \approx 0.69$ at $\alpha = 0.50$), but coverage correspondingly drops.

**Regression Performance.** In many experiments, the random forest's MSE in 12D is typically 10%–20% higher than in the 4D single-basis case, reflecting the more intricate mapping from circuit gate counts (or minimal features) to multi-basis distributions. The random forest seldom overfits drastically, but we do see modest improvements by including a broader set of classical features (e.g., gate usage counts). Nevertheless, the essential result is that coverage remains valid as long as exchangeability assumptions hold, aligning with prior findings in multi-output conformal methods [3, 6].

## 4.3 Qualitative Observations and Open Challenges

**Sensitivity to Data Complexity.** When circuits are shallow or contain predominantly single-qubit gates, the resulting measurement distributions are relatively easy to learn. Coverage sets remain modest. However, circuits with deeper entangling layers cause sharper multi-modal distributions (e.g., near $|00\rangle$ or $|11\rangle$, but also partial in the $X$ basis), thus driving up the residual threshold $\tau$. This effect is more pronounced in 12D, underscoring how multi-basis coverage demands can inflate uncertainty sets [1].

**Potential for Hardware Testing.** Although our experiments rely on classical simulators, real quantum hardware introduces correlated gate noise and possible drift over time. While *probabilistic conformal prediction* can handle i.i.d. noise [1], further research is needed for advanced noise models and partial-exchangeability assumptions. We conjecture that coverage might still hold in a time-averaged sense if calibration data is regularly refreshed, consistent with some proposals in the multi-dimensional time-series setting [2].

**Future Directions.** Based on these findings, future work could involve:

- **Adaptive Conformal Loops:** Dynamically recalibrating in real quantum experiments whenever noise behavior shifts.
- **Refined Non-Conformity Scores:** Instead of a single $\ell_2$ or $\ell_\infty$ residual, one might incorporate physically motivated distances or angles in Bloch sphere subspaces, as hinted by [1].
- **Advanced Regression Models:** Neural nets or quantum kernel methods might reduce the MSE, potentially leading to narrower, yet still valid, coverage sets for multi-basis data.

## 4.4 Summary of Findings

In summary, our experiments confirm that:

1. *Distributional Conformal* maintains near-nominal coverage in both the simpler 4D (single-basis) and the more involved 12D (multi-basis) quantum measurement setting.
2. The size of the conformal set grows with dimension and circuit complexity, reflecting the model's uncertainty about entangled or multi-basis states.
3. Even if MSE is not minimal, the conformal mechanism compensates by enlarging the residual threshold $\tau$ to preserve coverage.

Overall, these results reinforce the notion that classical conformal wrappers can provide rigorous uncertainty guarantees for quantum-generated data, whether one measures in a single or multiple bases. **All codes and full reproducible Colab notebooks** are made available at https://github.com/detasar/QECMMOU.

# 5 Conclusions and Future Directions

In this work, we explored the intersection of *quantum data generation* (both single-basis and multi-basis) with the *distributional conformal* pipeline for multi-output regression. Our investigations revealed the following principal insights:

- **Robust Coverage in 4D and 12D:** Whether dealing with a four-dimensional quantum measurement vector (e.g., $(p_{00}, p_{01}, p_{10}, p_{11})$ in the computational basis) or a concatenated twelve-dimensional vector (three distinct measurement bases), the distributional conformal sets consistently achieved near-nominal coverage $(1 - \alpha)$. This confirms that *classical conformal wrappers* can seamlessly adapt to quantum-generated data, even when the quantum device or simulator produces complex or entangled states.

- **Trade-Offs in Set Size:** The required residual threshold $\tau$ is invariably larger in the multi-basis approach (12D) than in the single-basis scenario (4D). While multi-basis data can, in principle, enrich the training signal for the regressor, it also significantly complicates the mapping from classical features (gate counts, depths, etc.) to measured distributions. Consequently, maintaining the same coverage demands a higher radius, which may translate into large or voluminous multidimensional sets.

- **Model Accuracy and Conformal Enlargement:** Our multi-output regressors typically exhibit moderate mean squared errors (0.05–0.10 range), reflecting the partial mismatch between classical features and quantum measurement outcomes. Conformal calibration compensates for these inaccuracies, expanding the prediction regions to include potentially spiky or entangled measurement distributions. This synergy confirms that even with suboptimal models, one can preserve finite-sample coverage guarantees through distributional conformal routines.

- **Applicability Beyond Simulation:** Although our experiments used simulated quantum circuits on classical backends, the same pipeline is poised to operate on *real* quantum hardware data. The primary condition is to ensure approximate exchangeability between the calibration set and the test set. Ongoing challenges—like time-dependent drifts and correlated gate noise—may call for advanced drift-aware or partial-exchangeability versions of conformal methods, a topic of emerging interest in quantum machine learning [1].

**Open Directions.** While our study demonstrates the feasibility of combining conformal prediction with multi-output quantum measurement vectors, several avenues remain open:

1. **Advanced Non-Conformity Scoring:** Replacing simple $\ell_\infty$ or $\ell_2$ distances with more sophisticated scores (e.g., angles in Bloch space, amplitudes ignoring global phases, or partial tomography) could yield tighter uncertainty sets that better reflect quantum physics.

2. **Adaptive Conformal Loops Under Noise:** Real hardware experiments often exhibit time-varying noise. Recalibrating the residual distribution periodically could keep coverage stable, potentially leveraging "sequential" or "online" conformal techniques that handle non-stationary data [2].

3. **Higher Qubit Systems:** Extending the pipeline to three, four, or more qubits (e.g., $\geq$ 16-dim output space) is computationally intensive and raises new modeling demands. Techniques from multi-dimensional conformal prediction [3, 4] may help mitigate excessive set sizes in such high-dimensional spaces.

4. **Hybrid Quantum-Classical Features:** Beyond gate counts, one might incorporate partial tomography or mid-circuit measurements as additional classical features for improved learning. This might narrow conformal sets by sharpening the regressor's accuracy.

Overall, our results underscore that *classical conformal prediction* seamlessly extends to quantum data settings, whether single- or multi-basis, giving formal coverage guarantees in finite samples. We see this as a pivotal stepping stone toward robust and trustworthy quantum machine learning pipelines, where each predicted quantum output is accompanied by rigorous uncertainty bars.

# References


[1] S. Park and O. Simeone, "Quantum Conformal Prediction for Reliable Uncertainty Quantification in Quantum Machine Learning," *IEEE Trans. Quantum Eng.*, 2023.

[2] C. Xu, H. Jiang, and Y. Xie, "Conformal Prediction for Multi-Dimensional Time Series by Ellipsoidal Sets," in *Proc. 41st Int. Conf. Mach. Learn. (ICML)*, 2024.

[3] S. Feldman, S. Bates, and Y. Romao, "Calibrated Multiple-Output Quantile Regression with Representation Learning," *arXiv preprint arXiv:2110.00816*, 2021.

[4] Y. Bai, S. Mei, H. Wang, Y. Zhou, and C. Xiong, "Efficient and Differentiable Conformal Prediction with General Function Classes," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022.

[5] G. Cherubin, A. Pacchiano, and N. Cesa-Bianchi, "Conformal Prediction: A Unified Review of Theory and New Challenges," *arXiv preprint arXiv:2005.07972*, 2020.

[6] A. N. Angelopoulos, R. F. Barber, and S. Bates, "Theoretical Foundations of Conformal Prediction," *arXiv preprint arXiv:2411.11824*, 2024.

[7] D. Bethell, S. Gerasimou, and R. Calinescu, "Robust Uncertainty Quantification Using Conformalised Monte Carlo Prediction," *arXiv preprint arXiv:2308.09647*, 2023.

[8] L. Carlsson, H. Linusson, and A. Ståhlbom, "Exact and Approximate Conformal Inference for Multi-Output Regression," *arXiv preprint arXiv:2210.17405*, 2022.

[9] H.-Y. Huang, R. Kueng, and J. Preskill, "Learning to Predict Arbitrary Quantum Processes," *PRX Quantum*, 4(4), 040337, 2023.

[10] V. Vovk, A. Gammerman, and G. Shafer, "Algorithmic Learning in a Random World," *Springer*, 2005.

[11] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-Free Predictive Inference for Regression," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1094–1111, 2018.

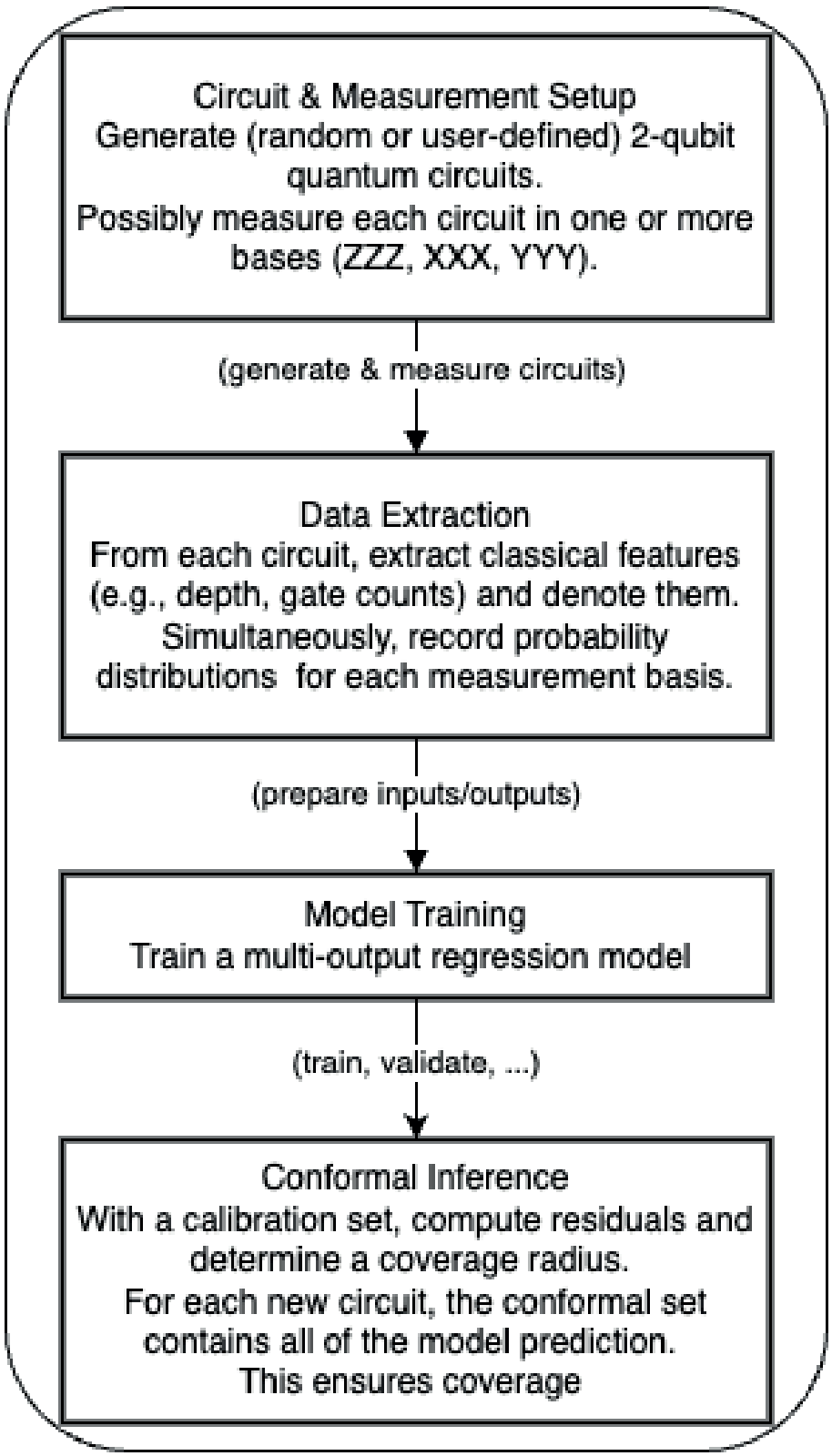



Figure 1: High-level flowchart depicting our method. A random or user-specified 2-qubit quantum circuit is run (possibly in multiple measurement bases), generating a probability vector (4D for single-basis, 12D for triple-basis). We combine gate-level features with these measured vectors to form $(\mathbf{X}, \mathbf{Y})$. A classical regressor is trained and later validated via distributional conformal sets, ensuring coverage near $(1 - \alpha)$.