

Adaptive Gated Mixtures of Experts for Kähler Potential Approximation: Metric Fidelity and Out-of-Distribution Evaluation

Som Subhro Nath

nathsomsubhro7@gmail.com

Independent Researcher <https://orcid.org/0009-0007-6842-0019>

Research Article

Keywords: Kähler potential, Hermitian metric, mixture of experts, automatic differentiation, out-of-distribution generalisation, metric positivity

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11459769/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Accurate approximation of a Kähler potential does not necessarily yield an accurate induced metric, since metric reconstruction depends on second derivatives. This study evaluates whether gated mixture-of-experts architectures improve potential and metric approximation relative to conventional and ungated neural baselines. Four architectures—a multilayer perceptron, a gated mixture of nonlinear experts, a uniformly averaged expert ensemble, and a gated model with linear experts—were compared on a synthetic Kähler potential with an analytically specified metric. The experiment comprised eight independently sampled datasets and three initialisation seeds per architecture. Performance was assessed using potential and metric mean squared errors on held-out and outer-region out-of-distribution (OOD) samples, together with sampled metric-positivity diagnostics. The full gated model achieved the lowest mean errors on both test and OOD sets. Relative to the multilayer perceptron, its mean metric MSE was lower by approximately 51.1% on the test set and 40.5% on OOD samples. It also exhibited the lowest observed fraction of non-positive metric eigenvalues among the evaluated models. However, errors increased substantially under OOD evaluation, and the gate diagnostics did not establish distinct expert specialisation. These results support gated expert aggregation as a promising approach for derivative-sensitive approximation of the prescribed synthetic potential. They do not establish global metric positivity, Ricci-flatness, or generalisation to other Kähler geometries.

1. Introduction

Machine-learning methods have been investigated as numerical approximators for geometric quantities arising in complex differential geometry. In particular, neural networks have been applied to the approximation of Kähler metrics on Calabi–Yau manifolds, where analytical expressions for Ricci-flat metrics are generally unavailable. Ashmore et al. [1] demonstrated that machine-learning models can approximate numerical Calabi–Yau metrics obtained through established computational procedures, while Jejjala et al. [2] investigated neural-network approximations on several Calabi–Yau geometries and evaluated the consistency of the resulting metrics. Subsequent work has extended the use of learning-based methods to the approximation of Kähler potentials and related geometric quantities [3]. These studies establish that neural approximation is a viable computational approach to certain geometric problems. They do not, however, imply that low prediction error alone is sufficient to establish the geometric validity of a learned metric.

A Kähler metric is locally determined by a real-valued potential K , through the complex-Hessian relation

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial z_i \partial \bar{z}_j}$$

Consequently, errors in a learned potential can propagate into its second derivatives and affect the resulting metric. A model with low potential error may therefore still exhibit appreciable metric error, particularly when derivative estimates are sensitive to approximation errors or when evaluation occurs

outside the training distribution. Potential approximation and induced-metric approximation should thus be treated as distinct evaluation objectives. Moreover, accurate approximation of the complex Hessian does not by itself establish positive-definiteness throughout the domain, and neither low metric error nor positive-definiteness alone establishes Ricci-flatness. These distinctions are essential when evaluating learning-based approximations of geometric objects.

One possible approach to improving approximation is to combine multiple expert functions through an input-dependent gating mechanism. In a mixture-of-experts model, the output is formed by weighting expert predictions according to a gate conditioned on the input. Unlike uniform aggregation, adaptive gating allows the contribution of each expert to vary across the input domain. This architecture provides a natural setting in which to test whether input-dependent aggregation improves approximation performance relative to a single neural network and to a fixed aggregation of the same expert family. However, the presence of a gating network does not itself establish interpretability, fuzzy-rule semantics, or improved generalisation. Such properties require independent definitions and empirical evidence. Research on neuro-fuzzy ensembles likewise identifies interpretability as an issue that must be examined rather than assumed from the use of hybrid or ensemble architectures [4].

The present study examines these questions on a controlled synthetic benchmark. The target is a polynomial Kähler potential expressed through symmetry-invariant variables. Its analytical form permits the corresponding complex-Hessian metric to be evaluated directly and checked against automatic differentiation. This construction provides a known reference against which both potential and metric predictions can be assessed. It also permits evaluation on held-out samples and on a separately specified out-of-distribution (OOD) region. The benchmark is deliberately limited: it is designed to compare approximation behaviour under controlled conditions, not to recover an unknown Ricci-flat metric, solve the Calabi conjecture numerically, or establish performance on general Kähler manifolds.

Four model configurations are evaluated: an invariant multilayer perceptron (MLP), a gated mixture of nonlinear experts, a uniform aggregation of the nonlinear experts, and a gated model with linear experts. *The comparison is organised around three research questions. First, does the gated nonlinear-expert model reduce potential and induced-metric errors relative to the invariant MLP? Second, does adaptive gating provide an advantage over uniform aggregation of the same nonlinear experts? Third, how does the linear-expert configuration compare with the nonlinear-expert model under in-distribution and OOD evaluation?* The experiments use eight independently sampled datasets and three random initialisation seeds per dataset. The analysis considers potential mean squared error, induced-metric mean squared error, and sampled metric-positivity diagnostics.

The contribution of this work is an empirical comparison of gated expert aggregation for potential and induced-metric approximation on a specified synthetic geometric target. The study examines whether observed gains persist across independently sampled datasets, whether they extend to OOD evaluation, and how they compare with fixed expert aggregation and a linear-expert configuration. The resulting evidence is interpreted at the level supported by the experimental design. In particular, conclusions are

restricted to the investigated target and sampling protocol; they are not taken as evidence of universal superiority across geometries. Likewise, sampled positivity diagnostics are treated as finite-sample numerical checks, not as guarantees of positive-definiteness over a continuous domain. By maintaining these distinctions, the study separates empirical approximation performance from stronger claims about geometric validity and theoretical generalisation.

2. Literature Review

2.1. Machine learning for Kähler and Calabi–Yau metrics

Machine learning has been applied to the numerical approximation of Kähler metrics, particularly in the context of Calabi–Yau geometry. Ashmore et al. [1] combined machine learning with numerical procedures based on Donaldson's algorithm to approximate Calabi–Yau metrics and associated quantities. Their results demonstrated that a learned approximation can reduce computational cost relative to the numerical procedure used to generate the reference metric. Jejjala et al. [2] investigated neural-network approximations for several Calabi–Yau geometries, assessing metric consistency, Ricci-flatness-related quantities, and discrete-symmetry behaviour. These studies established neural approximation as a relevant computational tool for geometric problems in which analytical metrics are generally unavailable.

Ashmore et al. [3] extended machine-learning approximations of Calabi–Yau metrics to more accurate numerical reference metrics, demonstrating that a Kähler potential could be predicted from a relatively small training sample. Ek et al. [5] investigated Grassmannian learning in conjunction with Donaldson's algorithm, using optimisation on the Grassmannian to identify efficient subspaces for metric construction. Mirjanić and Mishra [6] examined extrinsic symmetries of Calabi–Yau hypersurfaces and showed how such symmetries can yield compact representations of particular Ricci-flat metrics.

These studies should not be treated as interchangeable demonstrations of a single task. Learning a potential from numerical samples, approximating a metric tensor, and constructing a metric satisfying Ricci-flatness constraints involve different objectives and validation requirements. The present work addresses a narrower problem: *approximation of an analytically specified synthetic potential and the complex-Hessian metric induced by that potential*. It does not attempt to solve the general problem of constructing Ricci-flat metrics on Calabi–Yau manifolds.

2.2. Symmetry-aware representation and geometric validation

Symmetry can reduce the complexity of geometric learning when it is incorporated into the representation or enforced during training. Jejjala et al. [2] examined whether neural approximations reproduce the discrete symmetries expected of the underlying geometries. Mirjanić and Mishra [6] further demonstrated that additional symmetries can permit compact representations in particular

Calabi–Yau settings. These results motivate the use of invariant representations, but they do not imply that all architectures using invariant inputs have equivalent approximation properties.

In the present benchmark, the target potential is expressed in prescribed invariant coordinates. The models therefore operate on a representation that already incorporates the chosen invariance. The experiment tests approximation performance under that representation; it does not test whether a model can discover the invariance independently.

A second distinction concerns geometric validity. A Kähler metric is locally represented through the complex Hessian of a potential, but numerical approximation of that Hessian does not guarantee positive-definiteness over the domain. Ricci-flatness is a stronger condition and cannot be inferred from low elementwise metric error. Accordingly, the present study treats potential error, induced-metric error, and sampled metric positivity as separate outcomes. This is a diagnostic framework for the specified benchmark, not a substitute for the geometric validation required in Calabi–Yau metric construction.

2.3. Adaptive mixtures of experts

Adaptive mixtures of experts are an established neural modelling strategy. Jacobs et al. [7] introduced a learning procedure in which a gating mechanism allocates inputs to specialist networks, allowing different experts to address different portions of a learning problem. This architecture provides the conceptual basis for comparing input-dependent aggregation with fixed aggregation.

In a gated mixture of experts, the output is commonly expressed as

$$\widehat{K}(x) = \sum_{j=1}^M w_j(x) f_j(x), \quad \sum_{j=1}^M w_j(x) = 1,$$

where f_j denotes an expert and $w_j(x)$ is its input-dependent weight. In uniform aggregation, the weights are fixed at $1/M$. Adaptive gating introduces additional trainable parameters and permits the relative contribution of experts to vary with the input. Whether this flexibility improves generalisation is an empirical question rather than a consequence of the model definition.

The distinction between a gated expert model and a conventional multilayer perceptron is not limited to parameter count. The two configurations differ in functional decomposition and in how their outputs are combined. A comparison against an MLP can establish which configuration performs better under the tested conditions, but cannot isolate the contribution of gating. A uniform-expert baseline provides a more targeted comparison, although differences in parameterisation and optimisation must still be considered. Likewise, a comparison between linear and nonlinear experts is inconclusive about the effect of expert nonlinearity when the models have substantially different capacities.

These considerations motivate an ablation design that includes an invariant MLP, a gated nonlinear-expert model, uniform aggregation of the same nonlinear experts, and a gated linear-expert configuration. The objective is to characterise performance differences across these configurations rather than to establish a universal ordering of model families.

2.4. Neuro-fuzzy systems and interpretability

Neuro-fuzzy systems combine neural learning with fuzzy inference mechanisms, often using membership functions and fuzzy rules to represent relationships between inputs and outputs. Ouifak and Idri [4] reviewed neuro-fuzzy ensembles across application domains and discussed the relationship between ensemble performance and interpretability. Their subsequent review [8] examined fuzzy logic as a means of improving the interpretability and explainability of machine-learning models. These reviews make an important methodological distinction: the use of a hybrid architecture does not, by itself, demonstrate that its predictions can be explained through meaningful rules.

A softmax gating network produces normalised weights over experts. Those weights describe the model's aggregation mechanism, but they do not automatically correspond to fuzzy membership functions, linguistic variables, or a rule base. Consequently, a model employing ordinary softmax gating should not be labelled neuro-fuzzy without an explicit fuzzy-inference formulation.

This distinction is relevant to the present study because its principal evaluation concerns predictive performance rather than rule interpretability. Gate entropy and expert-output diagnostics can characterise how the model combines its experts, but they are not sufficient to establish interpretable specialisation. Unless the implementation contains a mathematically specified fuzzy mechanism, the model is more accurately described as a gated mixture of experts.

2.5. Derivative-sensitive error and out-of-distribution evaluation

When a predicted potential is differentiated to obtain a metric, function approximation error and derivative error are related but distinct. For a Kähler potential K , the associated metric is given locally by

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial z_i \partial \bar{z}_j}.$$

Small mean squared error in the values of K does not guarantee small error in its second derivatives. Evaluation based solely on potential values can therefore fail to identify models whose predictions yield inaccurate metric components.

The present work reports potential and induced-metric errors separately and includes evaluation on a specified out-of-distribution region. This provides a direct comparison of approximation behaviour under the selected training and evaluation distributions. The OOD protocol does not, however, represent a test of generalisation across distinct Kähler manifolds or different topologies.

Sampled metric positivity provides an additional diagnostic. Negative eigenvalues at evaluated points indicate that the corresponding numerical metric matrices are not positive-definite at those points.

Conversely, observing positive eigenvalues at finitely many sampled points cannot establish positive-definiteness throughout a continuous domain. Nor does positivity establish Ricci-flatness. The interpretation of these diagnostics must therefore remain separate from claims about geometric validity.

2.6. Knowledge gap and research positioning

The literature confirms that machine-learning approaches to Kähler and Calabi–Yau metrics are established [1–6], that adaptive mixtures of experts are a longstanding modelling approach [7, 9, 10], and that fuzzy interpretability requires an explicit methodological basis [4, 8].

The narrower research question is whether adaptive gating of nonlinear experts improves both potential approximation and induced complex-Hessian metric accuracy on a controlled, symmetry-invariant synthetic target, and whether any observed advantage persists under the specified OOD protocol.

Three specific issues motivate the experiments.

1. **Joint potential and metric evaluation.** Potential-value error alone does not characterise the accuracy of the induced metric. The study evaluates both quantities against an analytically specified target.
2. **Ablation of adaptive aggregation.** Comparing a gated nonlinear-expert model with an invariant MLP and a uniform-expert baseline provides evidence about the relative performance of the full adaptive configuration. The comparison with linear experts adds information about architecture, but parameter-count differences limit causal interpretation.
3. **OOD error and sampled positivity diagnostics.** These evaluations test whether the observed error differences extend beyond the held-out sample and whether sampled metric matrices exhibit negative eigenvalues. They do not establish global geometric validity or generalisation across distinct geometries.

The study addresses the following research questions:

- **RQ1:** Does the gated nonlinear-expert model achieve lower potential and induced-metric errors than the invariant MLP on the specified benchmark?
- **RQ2:** Does adaptive expert aggregation outperform uniform aggregation of the same nonlinear experts?
- **RQ3:** How does the gated linear-expert configuration compare with the nonlinear-expert model, given the difference in parameter count?
- **RQ4:** How do the configurations differ in OOD approximation error and sampled metric-positivity diagnostics?

The contribution is therefore an ablation-based empirical evaluation of gated expert aggregation for one analytically specified potential. It does not establish general superiority across Kähler potentials, demonstrate a new fuzzy-inference theory, or guarantee that the learned metric is positive-definite or Ricci-flat.

3. Methodologies

3.1. Problem Formulation

We consider the approximation of a prescribed Kähler potential K defined over a bounded domain in $\mathbb{C}^2$. Writing $z_j = x_j + iy_j$, the associated Hermitian metric is

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial z_i \partial \bar{z}_j}.$$

The objective is to determine whether gated expert aggregation improves approximation of the potential and its induced metric relative to conventional and ungated neural baselines. The target is a synthetic potential with a known analytic metric; it is not an unknown Ricci-flat Calabi–Yau metric. This controlled setting permits direct evaluation of potential error, metric error, and sampled metric positivity.

The target potential is defined through the invariants $u = |z_1|^2$, $v = |z_2|^2$, $s = u + v$, and $p = uv$:

$$K(s, p) = s + 0.12s^2 + 0.08p + 0.015s^3 + 0.02sp.$$

Using invariant inputs reduces the learning problem to two real variables while retaining a nontrivial dependence on the underlying complex coordinates.

3.2. Model Architectures

Four architectures were evaluated under a common regression protocol:

1. **Multilayer perceptron (MLP)**: a two-hidden-layer network with 48 tanh units per layer.
2. **Gated mixture of experts (G-MoE)**: a softmax gating network with one 16-unit tanh hidden layer and four expert networks, each comprising a 150-unit tanh hidden layer and a scalar output.
3. **Uniform expert ensemble**: four experts of the same architecture as the G-MoE experts, combined by an arithmetic mean without a learned gate.
4. **Gated linear experts**: the same gating architecture as the G-MoE, with four linear experts.

All models receive (s, p) and predict a scalar approximation $\widehat{K}$. In the G-MoE, the output is

$$\widehat{K}(s, p) = \sum_{k=1}^4 \alpha_k(s, p) f_k(s, p), \quad \sum_{k=1}^4 \alpha_k(s, p) = 1,$$

Where α_k are softmax gate weights and f_k denotes the k -th expert. The gate is a learned neural gating function; no explicit fuzzy membership functions or fuzzy rule base are implemented. The comparison therefore concerns gated expert aggregation, rather than a formal neuro-fuzzy inference system. The architectures follow the general mixture-of-experts framework introduced by Jacobs et al. [7].

3.3. Data Generation and Evaluation Splits

Eight independently generated datasets were used. For each dataset, 1,500 training points and 300 in-distribution test points were sampled independently from $[-0.5, 0.5]^4$ in real coordinates. An additional set of 200 out-of-distribution (OOD) points was drawn from $[-0.75, 0.75]^4$, retaining points satisfying

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 > 0.70.$$

The OOD protocol therefore probes approximation in an outer region of the coordinate domain, rather than generalisation to a different target geometry. Sampling was controlled through deterministic seeds. Each model was trained using three initialisation seeds, yielding $8 \times 3 \times 4 = 96$ model runs.

3.4. Training Protocol

All models were trained to minimise mean squared error (MSE) between the predicted and target potentials. Optimisation used Adam with a learning rate of 10^{-3} , batch size 256, and 180 epochs. No validation-based early stopping or learning-rate scheduling was used. The metric was not included explicitly in the training objective; metric fidelity was assessed after training through differentiation of the learned potential. This distinction is important because small potential error does not necessarily imply small second-derivative error, as derivative-based evaluation is also central to machine-learning approaches to Kähler geometry [1–3].

3.5. Metric Reconstruction and Error Measures

The learned metric was obtained from the real Hessian of $\widehat{K}$, using the coordinate convention

$$g_{ij} = \frac{1}{4} \left(\frac{\partial^2 K}{\partial x_i \partial x_j} + \frac{\partial^2 K}{\partial y_i \partial y_j} + i \frac{\partial^2 K}{\partial x_i \partial y_j} - i \frac{\partial^2 K}{\partial y_i \partial x_j} \right)$$

The analytic target metric was implemented independently from the derivatives of the prescribed potential. Agreement between the analytic metric and the auto-differentiated metric implementation was checked on 32 randomly sampled points; the maximum discrepancy was approximately 2.4×10^{-7} .

Potential MSE was evaluated over the full test and OOD sets. Metric MSE was computed as the mean squared complex-entry error over 100 deterministic subsamples from each set. Metric positivity was assessed by computing eigenvalues of the Hermitian-symmetrised predicted metric on the same subsamples. The reported negative-eigenvalue fraction is consequently a sampled diagnostic, not a proof of global positive definiteness.

3.6. Statistical Comparison and Gate Diagnostics

Models were compared using matched dataset–initialisation runs. For each baseline, paired differences were calculated relative to the full gated model, then averaged across the three initialisation seeds within each dataset. Summary statistics were therefore based on eight dataset-level means. Approximate 95%

intervals were calculated as the mean difference ± 1.96 standard errors and are treated as descriptive, given the limited number of independently generated datasets.

For the gated model, mean expert weights, dominant-expert assignment frequencies, and mean gate entropy were recorded. These measures describe routing behaviour but do not, by themselves, establish meaningful expert specialisation or interpretability.

4. Results

4.1. Potential Approximation and Metric Fidelity

The full gated mixture-of-experts model achieved the lowest mean potential and metric errors among the four evaluated architectures on both the in-distribution test set and the OOD set. The results are summarised in Table 1. Metric errors were consistently higher on the OOD samples, indicating that approximation quality deteriorated in the outer coordinate region.

Table 1
Mean potential and metric MSE across eight datasets and three initialisation seeds.

<i>Model</i>	<i>Test potential MSE</i>	<i>Test metric MSE</i>	<i>OOD potential MSE</i>	<i>OOD metric MSE</i>
MLP	0.000020	0.010937	0.029493	0.589159
Full gated MoE	0.000010	0.005349	0.014524	0.350506
Uniform experts	0.000015	0.007814	0.018737	0.401086
Gated linear experts	0.001067	0.063967	0.071063	0.465458

Relative to the MLP, the full gated model reduced mean test metric MSE by approximately 51.1% and OOD metric MSE by 40.5%. Relative to the uniform expert ensemble, the corresponding reductions were approximately 31.5% and 12.6%. The gated model also attained lower potential MSE than both nonlinear baselines on the test and OOD sets.

Mean test-set potential MSE across the four model configurations. Error bars indicate the variability reported by the experiment.

Mean test-set metric MSE for the four architectures. The full gated model has the lowest mean error.

These findings indicate that, for the prescribed target and sampling protocol, learned expert aggregation improved approximation of the induced metric beyond the gains obtained from using multiple experts alone. The smaller improvement over the uniform ensemble on OOD metric error suggests that the benefit of gating was more limited outside the training region than within it.

Mean potential MSE on the outer-region OOD samples. Error bars indicate the reported variability across runs or datasets, as defined in the plotting code.

Mean metric MSE on OOD samples. Metric error increases relative to the in-distribution test set, while the full gated model retains the lowest mean error

4.2. Paired Dataset-Level Comparisons

To assess consistency across datasets, paired differences were calculated by subtracting each baseline's error from the corresponding full gated model's error. Negative differences therefore favour the gated model. Table 2 reports the dataset-level summaries.

Table 2
Paired differences in MSE: full gated MoE minus baseline.

Baseline	Metric	Mean difference	95% approximate interval	Datasets favouring MoE
MLP	Test potential	-0.000010	[- 0.000014, - 0.000007]	8/8
MLP	Test metric	-0.005588	[- 0.007611, - 0.003565]	8/8
MLP	OOD potential	-0.014969	[- 0.017727, - 0.012210]	8/8
MLP	OOD metric	-0.238653	[- 0.268203, - 0.209103]	8/8
Uniform experts	Test potential	-0.000005	[- 0.000008, - 0.000003]	7/8
Uniform experts	Test metric	-0.002465	[- 0.003630, - 0.001300]	7/8
Uniform experts	OOD potential	-0.004213	[- 0.005897, - 0.002529]	7/8
Uniform experts	OOD metric	-0.050580	[- 0.069508, - 0.031652]	8/8

The gated model outperformed the MLP on all eight dataset-level comparisons across all four outcomes. Against the uniform expert ensemble, it achieved lower OOD metric error on all eight datasets and lower errors on the other outcomes on seven of eight datasets.

The intervals are normal-approximation intervals based on eight dataset-level means. They are descriptive rather than definitive inferential evidence; the limited number of datasets constrains the strength of statistical conclusions.

The comparison with gated linear experts was less uniform. The full gated model achieved lower test metric error and test potential error on all eight datasets, while its OOD metric advantage was observed

on six of eight datasets. The OOD metric difference had an approximate interval spanning zero. Moreover, the gated linear model has substantially fewer parameters than the full model, so this comparison changes both expert expressiveness and model capacity. It should not be interpreted as an isolated test of gating.

4.3. Sampled Metric Positivity

The fraction of sampled points at which the predicted Hermitian-symmetrised metric had at least one non-positive eigenvalue differed across architectures.

Table 3
Fraction of sampled OOD points with a non-positive metric eigenvalue.

Model	Negative-eigenvalue fraction
MLP	15.96%
Full gated MoE	2.79%
Uniform experts	4.04%
Gated linear experts	13.50%

The full gated model exhibited the lowest observed fraction, followed by the uniform expert ensemble. This pattern is consistent with the lower metric approximation error reported above, but it does not establish a causal relationship between metric error and positivity.

The result is *limited* to the sampled evaluation points. It does not establish global positive definiteness, and the learned models were not trained with an explicit positivity constraint. Consequently, the observed reduction should be interpreted as an empirical property of this experiment rather than a geometric guarantee.

4.4. Expert Utilisation

Gate diagnostics showed high normalised entropy, approximately 0.96 on OOD samples, with similarly high values on the test set. Thus, the gating network generally retained substantial uncertainty across expert assignments rather than consistently selecting a single dominant expert.

These diagnostics do not provide strong evidence of distinct expert specialisation. Although the gated architecture improved predictive performance, the current results do not establish that its experts correspond to identifiable geometric regimes or that the gate is interpretable in a substantive sense.

Mean gate weight assigned to each expert in the full gated model. The relatively balanced weights do not, by themselves, establish expert specialisation.

4.5. Summary of Findings

Across the evaluated synthetic problem, the full gated mixture of experts achieved the lowest mean potential and metric errors and the lowest sampled negative-eigenvalue fraction. Its advantage over the MLP was consistent across all eight datasets, while its improvement over uniform expert averaging was smaller but generally consistent. Performance deteriorated under the specified OOD sampling protocol for every architecture.

Taken together, the results support the use of learned gating as a promising approximation strategy for this particular synthetic Kähler potential. They do not establish generalisation to other potentials, global metric positivity, Ricci-flatness, or superiority on Calabi–Yau geometries.

5. Discussion

The results indicate that learned gating can improve the approximation of a synthetic Kähler potential and its induced metric relative to a conventional multilayer perceptron and a uniformly averaged expert ensemble. The strongest advantage was observed against the MLP, while improvements over uniform expert averaging were smaller. This distinction suggests that the performance gain cannot be attributed solely to the presence of multiple experts; the learned aggregation mechanism may also contribute to the observed improvement.

A central finding is the difference between potential approximation and metric fidelity. Since the metric is obtained from second derivatives of the potential, small errors in the predicted potential do not necessarily imply accurate metric reconstruction. The lower metric MSE of the full gated model on both test and OOD samples therefore provides a more geometrically relevant comparison than potential MSE alone. This observation is consistent with the broader emphasis on derivative-sensitive evaluation in machine-learning approaches to Kähler and Calabi–Yau geometry [1–3].

The OOD results require a more cautious interpretation. Although the gated model retained an advantage over both the MLP and uniform expert ensemble, metric errors increased substantially outside the training region. The experiment consequently demonstrates improved approximation under a particular outer-region sampling protocol, rather than robust generalisation to arbitrary domains or unseen geometries. Further evaluation across different target potentials, coordinate domains, and sampling distributions would be needed to establish the extent of this advantage.

The sampled positivity results provide an additional, but limited, indication of geometric fidelity. The full gated model exhibited fewer non-positive metric eigenvalues than the comparison architectures on the evaluated OOD points. However, positivity was neither enforced during training nor established throughout the domain. A potential with low regression error can still yield an indefinite Hessian-derived metric. Future implementations could investigate positivity-aware parameterisations or explicit geometric regularisation, while measuring whether these constraints improve validity without compromising approximation accuracy.

The gate diagnostics also limit the interpretation of the architecture. High normalised gate entropy indicates that the model generally distributes weight across several experts. It does not establish meaningful specialisation, and the present results do not show that individual experts represent distinct geometric structures. Accordingly, the principal contribution supported by this experiment is empirical performance comparison, not an interpretability claim.

Some limitations should be acknowledged. First, the study uses one analytically prescribed synthetic potential in complex dimension two. Second, the metric is evaluated on a deterministic subsample of the test and OOD sets. Third, the positivity assessment is finite-sample and does not establish global positive definiteness. Fourth, the gated linear-expert baseline differs substantially in parameter count from the full model, limiting architectural attribution. Finally, the dataset-level comparisons involve only eight independently generated datasets; their approximate confidence intervals should not be treated as definitive evidence of broad statistical generality.

Overall, the findings support gated expert aggregation as a promising architecture for derivative-sensitive approximation of the prescribed potential. They do not establish superiority for general Kähler potentials, recovery of Ricci-flat metrics, or satisfaction of the global geometric constraints required in Calabi–Yau metric construction. Those questions remain open for future evaluation.

6. Conclusion and Future Work

This study evaluated a gated mixture-of-experts architecture for approximating a prescribed synthetic Kähler potential and its induced Hermitian metric. Across eight datasets and three initialisation seeds, the full gated model achieved lower mean potential and metric errors than a conventional multilayer perceptron and a uniformly averaged expert ensemble. It also exhibited the lowest sampled fraction of non-positive metric eigenvalues among the evaluated architectures.

The results suggest that learned expert aggregation can improve derivative-sensitive approximation in this controlled setting. However, the observed advantage does not establish generalisation to arbitrary Kähler potentials or guarantee global metric positivity. The study is limited to one synthetic target, a bounded sampling protocol, and finite-sample geometric diagnostics. Furthermore, the gate diagnostics do not establish interpretable expert specialisation.

Future work should extend the evaluation to multiple analytic potentials, higher-dimensional complex domains, and more demanding OOD distributions. Explicit positivity constraints and metric-aware training objectives should also be investigated. Finally, comparisons under better-controlled parameter budgets and systematic ablations would help isolate the contribution of gating from that of expert capacity.

In summary, the present work provides a controlled empirical assessment of gated expert aggregation for potential-based metric approximation. Its findings motivate further investigation while leaving broader geometric validity and generalisation as open questions.

Declarations

Conflict of Interests:

No conflict of interests were identified in relation to the content, methodology, or findings of this study. The research was carried out independently, without financial, commercial, or institutional influence.

Funding Information:

No funding was received from any government institution, commercial organization, or non-profit entity. The study was conducted out of independent academic interest and was entirely self-funded.

7.1. Research Involving Humans and/or Animals:

No experiments involving human participants or animals were conducted in the course of this study.

7.2. Informed Consent:

Informed consent was not applicable, as no human participants were involved in the study.

Author Contributions:

The author conceived and designed the study, developed the experimental framework, conducted the experiments and analysis, and wrote and revised the manuscript.

Acknowledgements

Sincere acknowledgement is extended to **Mr. Ashim Kumar Nath, Bose Institute**, whose constant encouragement and intellectual influence have been instrumental in sustaining a rigorous engagement with scientific inquiry. His emphasis on critical questioning, methodological discipline, and evidence-based reasoning has provided enduring motivation throughout the conception and development of this work.

References

1. Ashmore A, He Y-H, Ovrut BA (2020) Machine learning Calabi–Yau metrics. *Fortschr Phys* 68:2000068. <https://doi.org/10.1002/prop.202000068>
2. Jejjala V, Kaloni Mayorga Peña D, Mishra C (2022) Neural network approximations for Calabi–Yau metrics. *J High Energy Phys* 2022:105. [https://doi.org/10.1007/JHEP08\(2022\)105](https://doi.org/10.1007/JHEP08(2022)105)
3. Ashmore A, Calmon L, He Y-H, Ovrut BA (2022) Calabi–Yau metrics, energy functionals and machine-learning. *Int J Data Sci Math Sci* 1(1):49–61. <https://doi.org/10.1142/S2810939222500034>

4. Ouifak H, Idri A (2023) Application of neuro-fuzzy ensembles across domains: A systematic review of the two last decades (2000–2022). *Eng Appl Artif Intell* 124:106582. <https://doi.org/10.1016/j.engappai.2023.106582>
5. Ek CH, Kim O, Mishra C (2024) Calabi–Yau metrics through Grassmannian learning and Donaldson’s algorithm. *arXiv:2410.11284*. <https://arxiv.org/abs/2410.11284>
6. Mirjanić V, Mishra C (2024) Symbolic approximations to Ricci-flat metrics via extrinsic symmetries of Calabi–Yau hypersurfaces. *arXiv:2412.19778*. <https://arxiv.org/abs/2412.19778>
7. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE (1991) Adaptive mixtures of local experts. *Neural Comput* 3(1):79–87. <https://doi.org/10.1162/neco.1991.3.1.79>
8. Ouifak H, Idri A (2025) A comprehensive review of fuzzy logic based interpretability and explainability of machine learning techniques across domains. *Neurocomputing* 647:130602. <https://doi.org/10.1016/j.neucom.2025.130602>
9. Jordan MI, Jacobs RA (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comput* 6(2):181–214. <https://doi.org/10.1162/neco.1994.6.2.181>
10. Chen K, Xu L, Chi H (1999) Improved learning algorithms for mixture of experts in multiclass classification. *Neural Netw* 12(9):1229–1252. [https://doi.org/10.1016/S0893-6080\(99\)00043-X](https://doi.org/10.1016/S0893-6080(99)00043-X)
11. Code and Data Availability
12. The source code benchmark datasets, experimental outputs, and generated figures associated with this study are publicly available on Zenodo: <https://doi.org/10.5281/zenodo.23541089>. The repository provides the computational materials supporting the reported experiments and is intended to facilitate independent inspection and reproducibility

Figures

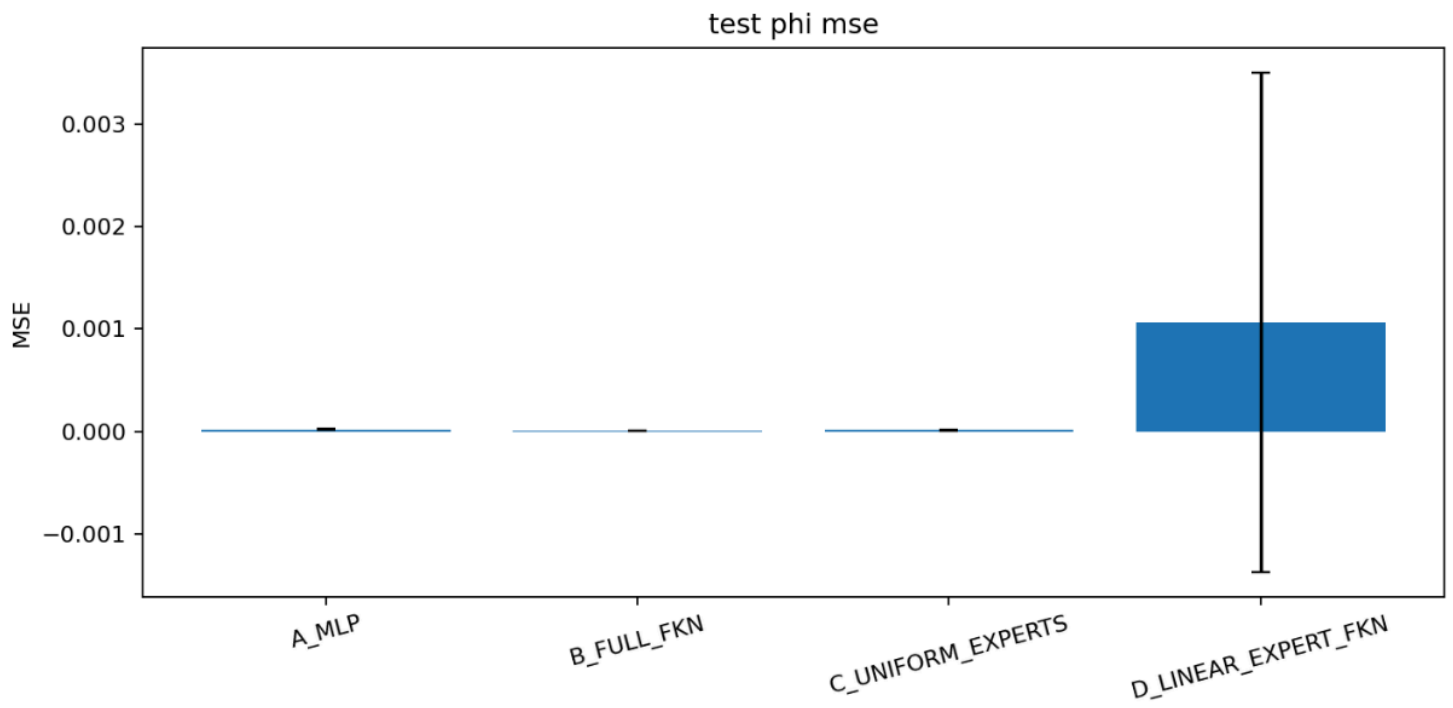


Figure 1

In-Distribution Potential Error

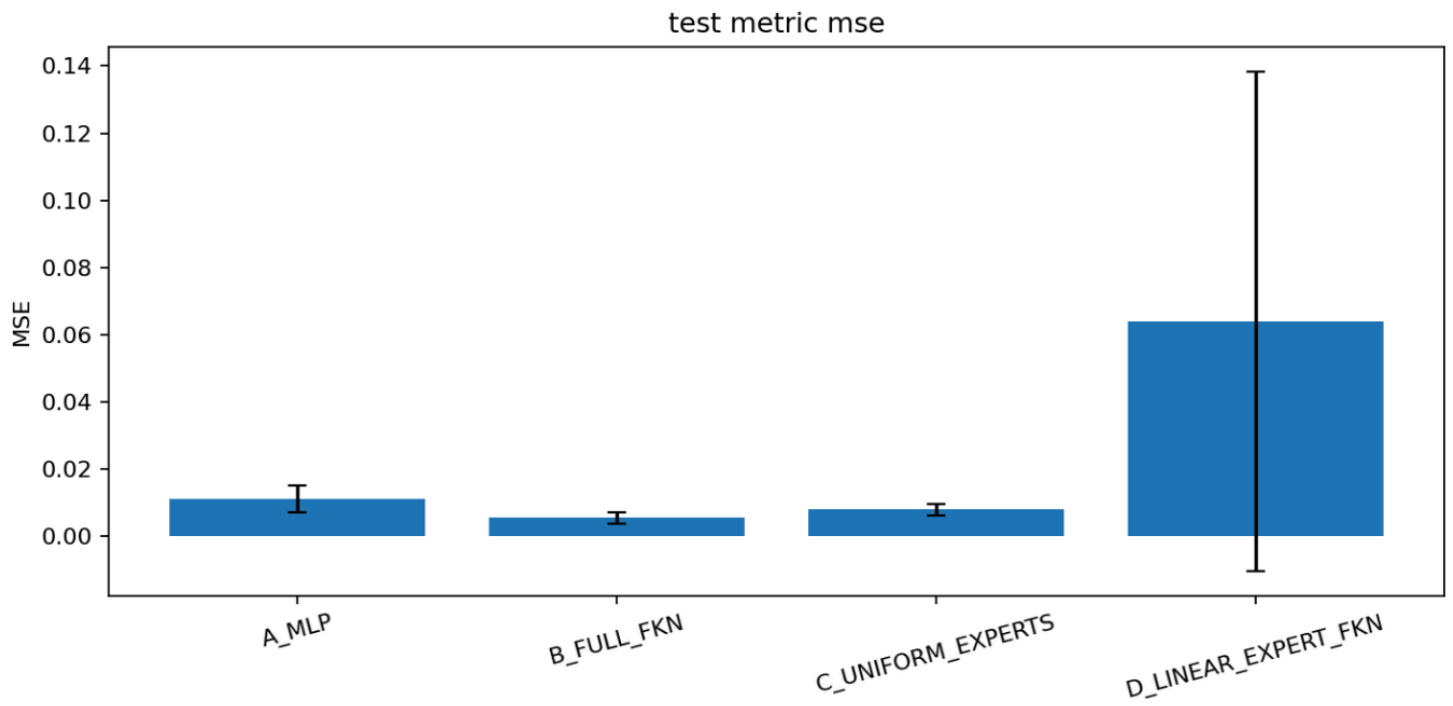


Figure 2

In-distribution Metric Error

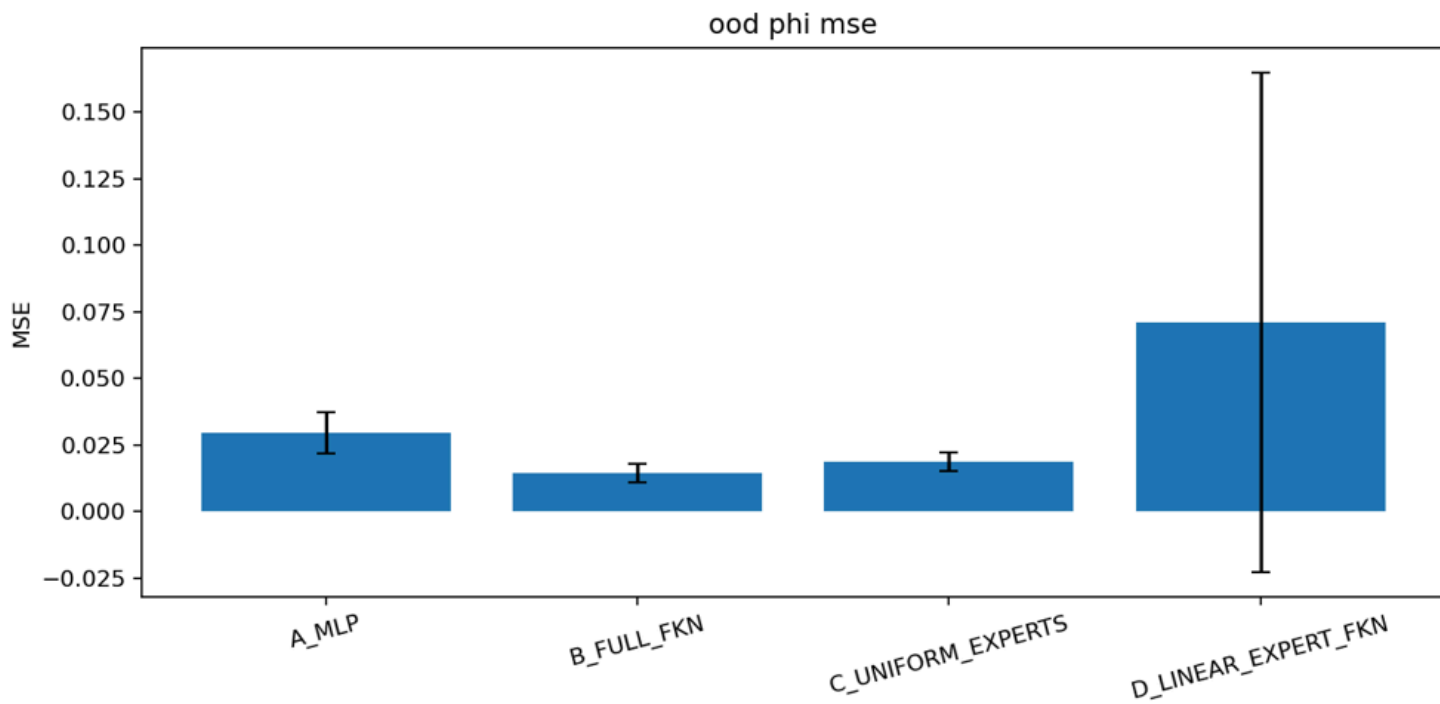


Figure 3

OOD potential error

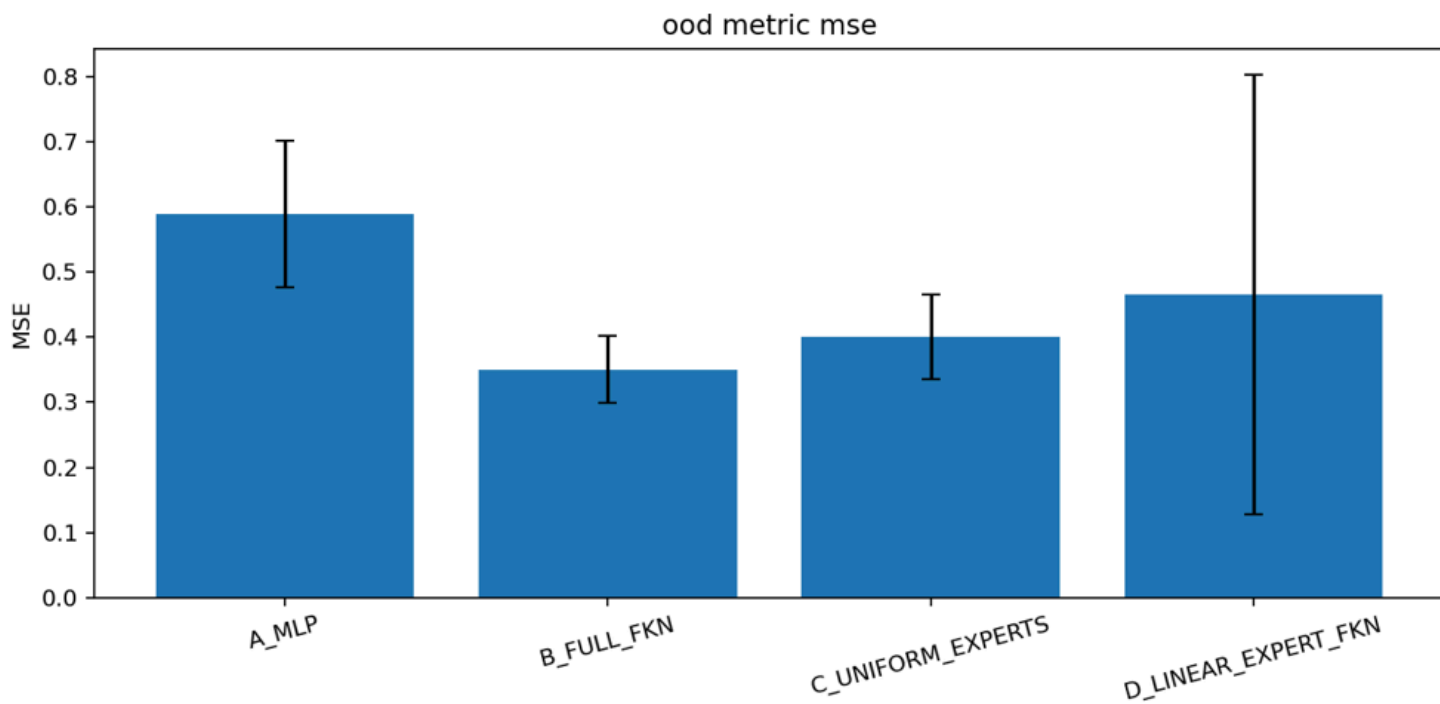


Figure 4

OOD metric error

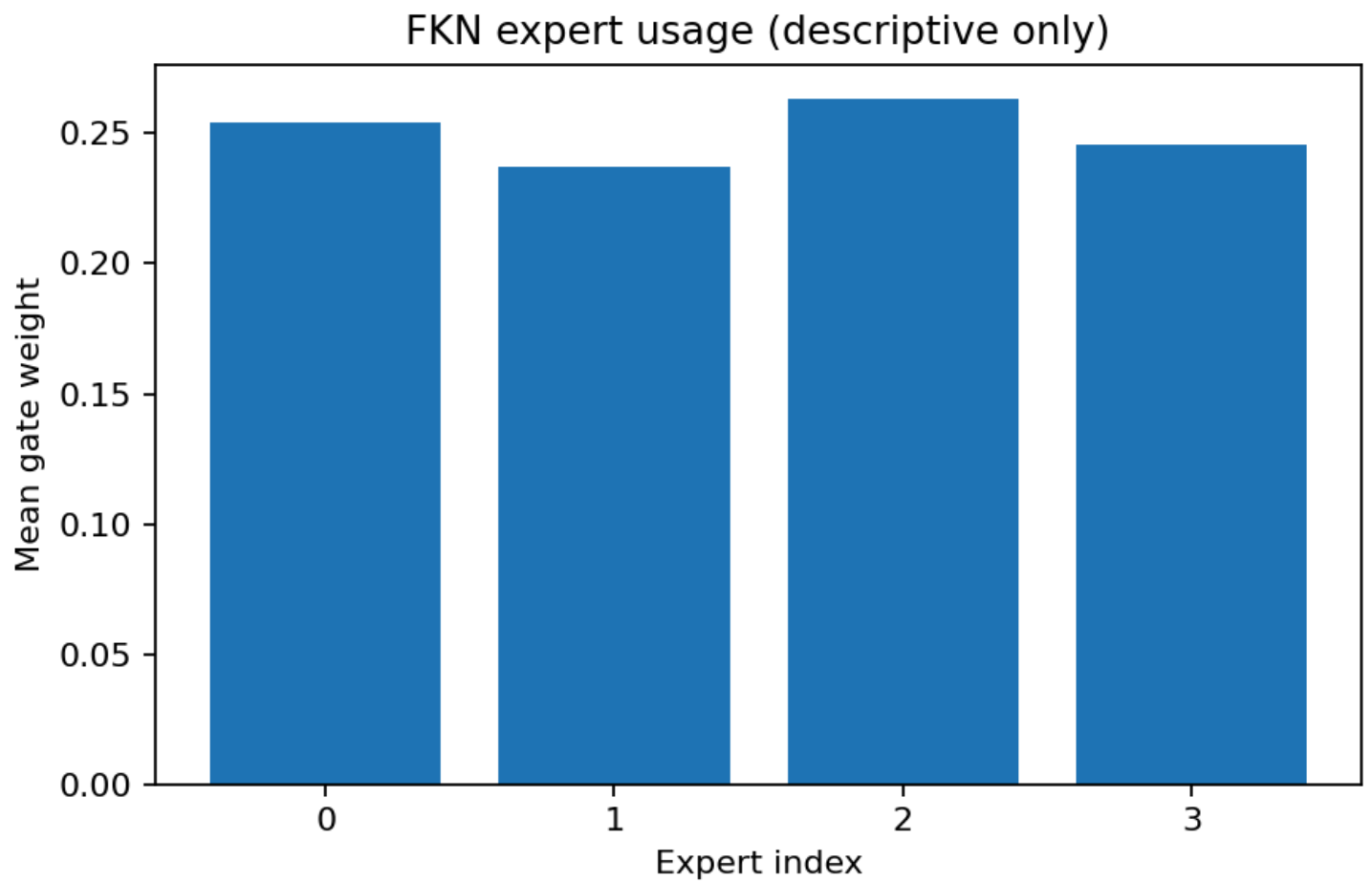


Figure 5

Expert Utilisation