

Initialization-Free Non-Linear Constrained Optimization Using a Bayesian Self-Supervised MLP

Mahamudul Hassan Siddique, Miftaur Rahman Zisan and Fahimul Haque

Department of Industrial and Production Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

hassansiddique632@gmail.com, miftaurr314@gmail.com, fahimulhaq2001@gmail.com

Abstract

Nonlinear constrained optimization requires a balance in the approach, exploration and minimization of objectives. In this paper, a Bayesian self-supervised multi-layer perceptron (MLP) model is presented, which focuses on the satisfaction of constraints as well as minimizing user reliance on initialisation. As opposed to classical solvers, there must be no initial guess, candidate solutions are drawn using standard normal distribution and refined. Bayesian Optimization is a two-fold approach that automates the process of supervising hyperparameters as well as provides intelligent search-space exploration to help the model to escape local optima and converge to viable near-global solutions. The framework integrates differentiable inequalities, equality and bound constraints within the loss to provide a strong feasibility to the loss. Training is also stabilized further with gradient clipping, regularization and input-parameter co-optimization. The benchmark assessments of the model in standard test functions and engineering design problems reveal that the model is always able to find feasible solutions with competitive or better objective values than IPOPT indicating that it is indeed a robust model as a feasibility-oriented surrogate solver.

Keywords

Bayesian Optimization, Nonlinear Optimization, Constrained Optimization, Self-Supervised Learning, Neural Networks, Multi-Layer Perceptron.

1. Introduction

Nonlinear constrained optimization is one of the key problems in science and engineering, in which decision variables are constrained by both multidimensional objective functions as well as by strict feasibility conditions. These issues are experienced in various areas of work like mechanical design, energy planning, structural engineering and process control. The traditional solvers, such as sequential quadratic programming and interior-point methods, are still very successful at convex or small-scale problems, but are severely challenged in the high-dimensional landscapes where multiple local optima and nonconvex constraints can be found. In addition, such techniques are commonly based on judiciously selected initial guesses, which may be hard to provide in practice, and tend to provide even slightly infeasible results following convergence.

In order to overcome these drawbacks, the present paper presents a Bayesian self-supervised neural architecture which is built on multi-layer perceptron (MLPs). Our model does not have an initial specification as in classical solvers, but rather candidate solutions are sampled using a standard normal distribution and improved by iteration. Constraint satisfaction is the primary priority and Bayesian Optimization also offers adaptive hyperparameter tuning and search space exploration. By adopting a fair exploration and exploitation scheduling, Bayesian Optimization can help the model escape local optima to get practical near global solutions.

1.1 Research Objectives

- 1.To create a single self-supervised neural network that self-learns to combine objective learning, constraint enforcement and exploration of solutions without initial guesses provided by the user.
- 2.To use Bayesian Optimization as a tool to tune hyperparameters as well as exploration of the search-space, to allow the model to successfully get out of local optima.
- 3.To establish a feasibility-first loss formulation with differentiable penalties on inequality, equality, and boundary constraints, to take the constraint satisfaction of nonlinear problems.
- 4.To test the suitability of the proposed solution on benchmark and engineering design problems we plan to compare the performance of this solution to the state-of-the-art solver like IPOPT, and focus more on feasibility, robustness, and flexibility.

2. Literature Review

During the last half a dozen years there has been an astonishing interest in a study of nonlinear constrained optimization. While methods like sequential quadratic programming, augmented Lagrangians based algorithms or interior point solvers have been the main workhorses in prior art, the current generation employs surrogates, Bayesian optimization (BO), neural networks, and such. These improvements are not only aimed at improving efficiency but also to deal with some open issues about reliance on manually selected starting guesses, falling into local minima and violations of strict feasibility constraints. One recurring theme of this new collection is the move from post-hoc constraint enforcement to constraint management enmeshed in, and from deterministic iterative methods to uncertainty-aware exploration policies.

A classic example of masking is the Surrogate-Assisted Global and Distributed Local Collaborative Optimization (SGDLCO) approach suggested by Liu et al. (2025). SGDLCO consists of two complementary modules per generation: a global surrogate-assisted search phase to explore the wide search space and a local surrogate-assisted optimization phase to strengthen candidate good areas. Notably, a three-layer adaptive selection scheme is devised to better compromise among feasibility, diversity and convergence. In contrast to prior surrogate-based methods, which were dominated by convergence rates and suffered from rarely mentioned constraint satisfaction and diversity, SGDLCO focuses especially on constraint satisfaction and diversity to discourage premature convergence. Competitive or best performance on expensive constrained problems in high-dimensional spaces, illustrating the benefit of global exploration within a nested local surrogate refinement scheme (Liu et al. 2025).

The local optima escape issue has been approached from different aspects. One of the biologically inspired approaches is the Hare Escape Optimization (HEO) algorithm introduced by Alsamee and Ramezani (2025). Inspired by predator-escape behavior in nature, HEO is a combination of long-distance jumps modeled by Lévy flights and directionally biased modifications that resemble evasive maneuvers. This enables the algorithm to avoid stagnation in narrow local basins without compromising feasibility. Thoroughly tested on benchmark engineering design challenges such as welded-beam and spring optimization, HEO consistently showed viable solutions with satisfactory objective performance. At a more general level, HEO illustrates how surrogate-driven optimization can be supplemented by the integration of metaheuristic escape mechanisms introducing stochasticity at strategic points (Alsamee and Ramezani 2025).

Bayesian optimization itself has been especially well suited to constrained scenarios because of its capacity for balancing exploration and exploitation using probabilistic surrogate models. Røstum et al. (2025) demonstrated this in a design case study for a bridge where both objective function and constraints were substituted with probabilistic surrogates. Using acquisition functions such as constrained expected improvement and constrained max-value entropy, the method converged rapidly to practical, high-fidelity designs within a few iterations. The study emphasized that uncertainty modeling is not only a worldwide exploration method but also a feasibility management mechanism: by identifying uncertainty within constraint projections, the algorithm avoids over-exploitation that may lead to infeasible solutions (Røstum et al. 2025).

The efficiency of BO frameworks can also be improved by combining them with pretrained models. Picard et al. (2024) introduced PFN-based BO using Prior-data Fitted Networks (PFN) - transformer networks trained on synthetic optimization tasks. Such networks allow BO to skip costly initial sampling phases and generalize across optimization task families. Although the work does not specifically focus on constrained environments, it showed that PFN-based

BO can be made constraint-aware by conditioning predictions based on feasibility indicators. This efficient, accurate, and PFN-BO indeed can be a good option to optimize engineering pipelines with high complexity (Picard et al. 2024).

The other extension is SDPOA introduced by Yang et al. (2025). Rather than having a predetermined size population, SDPOA tunes candidate pools based on surrogate estimates of objective quality and feasibility. The adaptive population part will make it less prone to prematurely converging and stay in the low-sample regions. Applied to computationally expensive constrained problems, SDPOA provided superior diversity and fewer constraint violations than classical evolutionary algorithms. Its design demonstrates how population dynamics and surrogates can collaborate to ensure feasibility and exploration (Yang et al. 2025).

At the interface between neural networks and optimization, Kilwein et al. (2023) recommended using neural feasibility surrogates. These classifiers project constraint satisfaction across the domain and are integrated in optimization loops to eliminate infeasible candidates early on. By separating feasibility evaluation from expensive objective calculations, the approach reduces computational effort significantly. The framework was used in energy system optimization problems and showed that neural feasibility surrogates could accelerate convergence without sacrificing constraint satisfaction (Kilwein et al. 2023). Gu et al. (2022) generalized the surrogate-assisted optimization to discrete multi-objective problems. Employing random-forest surrogates coupled with various adaptive ranking strategies, their technique efficiently handled constraints and diversity of solutions under limited evaluation budgets (Gu et al. 2022). Apart from the selection of surrogates, active learning of constraints is another important element of innovation. Lei et al. (2021) proposed that conventional Gaussian process surrogates could be brittle in high-dimensional or non-smooth problem spaces. They proposed replacing them with Bayesian multivariate adaptive regression splines (BMARS) and Bayesian additive regression trees (BART), which are able to fit more adaptively to complex landscapes. They confirmed in experiments on automated experimental design that adaptive surrogates improved robustness and decreased vulnerability to local minima (Lei et al. 2021). Khatamsaz et al. (2023) extended this work, applying entropy-based active learning to the exploration of feasibility boundaries in alloy design. Their tactic involved sampling around free boundaries in a sense and essentially generated areas of potential interest before committing to true objective optimization. This is an example of constraint-first exploration, where feasibility is predicted in advance and not enforced afterwards (Khatamsaz et al. 2023).

Boundary-dominant acquisition has also been formalized in the BE-CBO (Boundary Exploration Constrained Bayesian Optimization) approach of Tian et al. (2024). Unlike approaches that start with an initial feasible point, BE-CBO allows for optimization to begin from infeasible samples and then slowly converge towards the boundary of the feasible and infeasible regions. Since many optima are in or near the vicinity of the feasibility boundaries, this method increases the chance to find close-to-global feasible solutions. For engineering problems involving sparse and scattered feasible regions, BE-CBO is an important step forward (Tian et al. 2024). Concurrently, Li (2025) suggested a BO approach with active constraint learning by refining surrogates of constraints on the fly during optimization. By minimizing use of explicit constraint computation and considering possible feasible regions, the technique results in higher convergence rates, robustness on local minima rich landscapes (Li 2025).

Also complementary are multi-fidelity approaches including those of Khatouri et al. (2020) that use inexpensive low-fidelity models combined with high-fidelity analyses. They effectively optimize and maintain constraint satisfaction by trading feasibility information in different levels of loyalty. Paulson and Tsay (2024) give an overview of the likes of such frameworks in exploration systems, where they observe that adaptive acquisition and modular surrogate structures are key for preserving active exploration while preventing them from getting stuck to local optima.

Decomposition Propagation Since then, algorithmic improvements of BO itself have been developed. A constrained knowledge-gradient acquisition function (c-KG) was introduced by Chen et al. (2021) which generalizes BO to address several constraints by basing it on the direct measurement of the value of information from feasible exploration. Ungredda and Branke (2021) generalized these concepts to include information-theoretic measures for the purpose of exploring feasible and infeasible regions. Zhang et al. (2023) also showed that considering dependencies between objectives and constraints can make a significant difference in efficiency (empirically). Together, these computational advances validate the supporting of constrained BO and escalate its application to practice (Chen et al. 2021; Ungredda and Branke 2021; Zhang et al. 2023).

Integrating across these works, several patterns emerge. One, uncertainty-aware surrogates ranging from Gaussian processes through to PFNs and neural classifiers are pivotal in exploration-feasibility balance. Second, proactive constraint learning and boundary exploration become increasingly highlighted as means for speeding up viable solution discovery and preventing fruitless evaluations. Third, explicit escape mechanisms whether Lévy flights, stochastic populations, or uncertainty-driven acquisitions are essential for traversing nonconvex landscapes. With all this progress, there remain gaps: many techniques demand delicate hyperparameter tuning, hard feasibility guarantees remain brittle under noisy evaluations, and few frameworks combine all desired properties at once. The literature thus hinges on the fact that next-generation methods must include initialization-free sampling, constraint-first optimization, uncertainty-driven exploration, and local minima robustness, a path the present research seeks to advance.

3. Methods

3.1 Framework Overview:

A self-supervised neural network framework is proposed for solving non-linear constrained optimization problems. A multi-layer perceptron (MLP) is employed to map a learnable dummy input vector into the solution space. This dummy vector is optimized jointly with the network parameters through automatic differentiation, enabling exploration of the feasible region in a differentiable manner. Feasibility is enforced and convergence is guided by embedding the objective and constraints directly into the loss function. The entire system is trained end-to-end using gradient-based optimization. Flexibility in handling diverse non-linear objectives and constraint types is provided by this formulation, while scalability to high-dimensional problems is maintained (Figure 1).

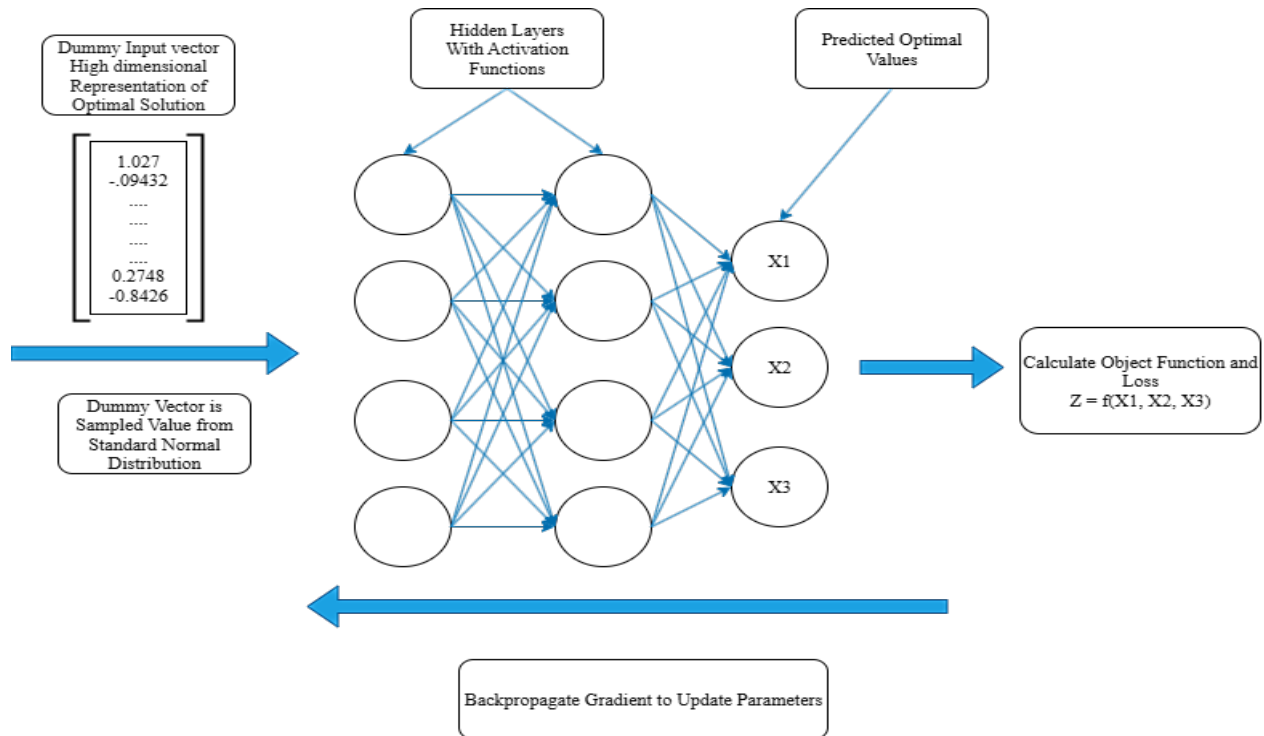


Figure 1. Proposed Framework

The whole framework can be seen in Figure 1. Unlike classical solvers, in which explicit feasibility checks are required at each step, the solution space is smoothly explored, and optimal feasible solutions are approached through the use of neural approximators. Adaptability is further enhanced by the integration of Bayesian Optimization for hyperparameter tuning, ensuring that convergence speed, accuracy, and constraint satisfaction are dynamically balanced. In this manner, objective-directed learning, constraints imposition and automated hyperparameter search are integrated into a single coherent pipeline for constrained optimization.

3.2 Loss Function Design:

The loss formulation is constructed by integrating three components: objective loss, constraint penalties, and regularization. To stabilize optimization, the objective function is shifted and squared:

$$L_{obj} = (f(x) + B)^2, \quad B \approx \min(\text{unconstrained } f(x))$$

so that non-negativity is ensured, and effective gradient propagation is maintained even when negative values are admitted by the objective. Constraint violations are penalized through differentiable penalty functions. Squared ReLU penalties are adopted for inequality constraints, squared residuals are employed for equality constraints, and interval-based penalties are used to capture bound violations. All constraint losses are aggregated into three terms, L_{ineq} , L_{eq} , L_{bound} , with weighting applied to control the strictness of feasibility enforcement. Finally, ℓ_2 regularization is introduced to impose a smoothness penalty on network weights, thereby preventing overfitting and excessive oscillations in the parameter space. The final objective is expressed as:

Equality Constraint Loss Function:

$$L_{eq} = (C(x))^2, \text{ Where } C(x) \text{ is a equality constraint}$$

$$\text{Inequality Constraint Loss Function: } L_{ineq} = (\text{ReLU}(C(x) - c_1))^2, \\ \text{Where } C(x) \leq c_1 \text{ is a inequality constraint}$$

Bound Constraint Loss Function:

$$L_{bound} = (\text{ReLU}(\mathbf{x} - \mathbf{u}) + \text{ReLU}(\mathbf{x} - \mathbf{l}))^2,$$

Where \mathbf{x} are the decision variables and \mathbf{u} upper bound and \mathbf{l} lower bound

Total Loss:

$$L_{total} = L_{obj} + \alpha L_{eq} + \beta L_{ineq} + \gamma L_{bound} + L_{reg} \text{ Where } \alpha, \beta, \gamma \text{ are the penalty weights}$$

Through this formulation, simultaneous reduction of the original objective, satisfaction of constraints, and maintenance of stable generalization are ensured. By embedding all problem requirements into the loss function, the network is trained as a direct surrogate solver, whereby feasible and near-optimal solutions are produced without the need for iterative feasibility checks. In this way, scalability to diverse optimization scenarios is achieved.

3.3 Training Stabilization:

To ensure robust and stable convergence, gradient clipping and joint optimization of inputs and weights are incorporated. During training, exploding gradients are mitigated by clipping the gradient norm to a predefined threshold τ , such that updates are kept bounded:

$$g_{clip} = g \cdot \min\left(1, \frac{\tau}{\|g\|_2}\right)$$

In this way, instability is prevented and smooth progress along the optimization trajectory is ensured. In addition, the dummy input vector, serving as a latent representation of the candidate solution, is optimized jointly with network parameters through automatic differentiation. By this mechanism, exploration of feasible solution spaces is enabled more effectively compared to fixed input mappings. Together with ℓ_2 regularization, optimization dynamics are stabilized, oscillatory updates are reduced, and convergence on non-linear constrained landscapes is facilitated. Through the combination of gradient clipping, weight decay, and differentiable input optimization, common pitfalls of instability are avoided during training while flexibility to adapt to a wide range of constrained optimization problems is maintained.

3.4 Hyperparameter Optimization:

The effectiveness of the proposed model is highly dependent on hyperparameter choices, including learning rate, weight decay, network size, penalty coefficients, and optimizer selection. To automate this process, Bayesian Optimization with a Tree-structured Parzen Estimator (TPE) sampler is employed, which efficiently balances exploration of uncertain configurations with exploitation of promising regions. At each iteration, the objective landscape of hyperparameters is approximated by the surrogate model, and the next candidate set for evaluation is selected by an acquisition function. In this way, optimal configurations are discovered efficiently under limited computational budgets. Specifically, the learning rate is tuned within $[10^{-8}, 10^{-1}]$, the number of neurons per layer is

varied within [8, 1024], penalty weights for constraint satisfaction are adjusted across several orders of magnitude, and multiple optimizers such as Adam, SGD, and RMSprop are considered. Through this approach, adaptation to different problem scales and complexities is achieved without manual tuning. By employing Bayesian Optimization, convergence speed is improved, the risk of local optima is reduced, and solution quality is enhanced compared to fixed hyperparameter settings. The proposed model is evaluated on several standard nonlinear constrained optimization benchmarks, and its performance is compared against IPOPT (CasADi). In each subsection, the problem formulation (objective and constraints) is presented, followed by a compact result table.

4. Results and Discussion

The proposed model is evaluated on several standard nonlinear constrained optimization benchmarks, and its performance is compared against IPOPT (CasADi). In each subsection, the problem formulation (objective and constraints) is presented, followed by a compact result table.

4.1 Branin Problem

$$\text{Minimize : } f(x_1, x_2) = a(x_2 - bx_1^2 + cx_1 - r)^2 + s(1 - t) \cos(x_1) + s$$

$$\text{Where } a = 1, b = 5.14\pi^2, c = 5\pi, r = 6, s = 10, t = 18\pi \text{ \& } x_1, x_2 \in [-5, 10]$$

Subject to:

$$\max (0, 0.5 - \sin(x_1 + x_2)^2) \leq 0$$

The results of the Branin problem is given in Table 1.

Table 1. Branin Problem Results

Method	Optimal Solution [x1,x2][x1, x2][x1,x2]	Objective Value	Constraint Viol.
Our Model	[3.2416, 3.5664]	2.3174	0
IPOPT	[3.2097, 3.5971]	2.3097	9.21×10^{-9}

4.2 Himmelblau's Problem

$$\text{Minimize : } f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

Subject to:

$$x_1^2 + x_2 \leq 4$$

$$x_1 + x_2^2 \leq 3$$

The results of the Himmelblau's problem is given in Table 2.

Table 2. Himmelblau's Problem Results

Method	Optimal Solution	Objective Value	Constraint Viol.
Our Model	[1.6906, 1.1340]	65.2978	0
IPOPT	[2.2109, -0.8883]	65.0000	9.69×10^{-9}

4.3 Constrained Rosenbrock

$$\text{Minimize : } f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Subject to

$$x_1^2 + x_2^2 - 2 \leq 0$$

$$x_1 + x_2 - 1 \leq 0$$

$$x_1, x_2 \in [-5, 5]$$

Table 3 shows the detailed result of the problem.

Table 3. Constraint Rosenbrock Problem Results

Method	Optimal Solution	Objective Value	Constraint Viol.
Our Model	[0.6187, 0.3813]	0.145617	0
IPOPT	[0.6188, 0.3812]	0.145607	1.0×10^{-8}

4.4 Pressure Vessel Design

$$\text{Minimize : } f(x_1, x_2, x_3, x_4) = 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3$$

Subject to:

$$\begin{aligned} 0.0193x_3 - x_1 &\leq 0 \\ 0.00954x_3 - x_2 &\leq 0 \\ \pi x_2^3x_4 + 43\pi x_3^3 - 1296000 &\leq 0 \\ x_1, x_2 \in [0.0625, 99], x_3, x_4 \in [10, 200] \end{aligned}$$

The comparison between the results is given in Table 4.

Table 4. Pressure Vessel Design Problem Results

Method	Optimal Solution	Objective Value	Constraint Viol.
Our Model	[0.19387, 0.09681, 10.0150, 10.0444]	38.0667	0
IPOPT	[0.19300, 0.09540, 10.0000, 10.0000]	37.5449	0

4.5 G04- Gear Train Design

$$\text{Minimize : } f(x_1, x_2, x_3, x_4, x_5) = 5.3578457x_3^2 + 0.8356891x_1x_5 + 37.293239x_1 - 40792.141$$

Subject to:

$$\begin{aligned} C_1 &= 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5 \\ C_2 &= 80.51249 + 0.0071317x_2x_3 + 0.002995x_1x_2 + 0.0021813x_3^2 \\ C_3 &= 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_4 \\ 85 \leq C_1 \leq 92, \quad 90 \leq C_2 \leq 110, \quad 20 \leq C_3 \leq 25 \\ x_1 \in [78, 102], \quad x_2 \in [33, 45], \quad x_3, x_4, x_5 \in [27, 45] \end{aligned}$$

The results of the Gear Train Design problem is given in Table 5.

Table 5. Gear Train Design Problem Results

Method	Optimal Solution	Objective Value	Constraint Viol.
Our Model	[78.28, 33.00, 32.68, 35.13, 39.48]	-29565.1602	0
IPOPT	[77.99, 32.99, 45.0, 26.99, 26.99]	-25273.6512	2.10×10^{-6}

4.6 G06- Global Optimization

$$\text{Minimize : } f(x_1, x_2) = (x_1 - 10)^3 + (x_2 - 20)^3$$

Subject to

$$\begin{aligned} -(x_1 - 5)^2 - (x_2 - 5)^2 + 100 &\leq 0 \\ (x_1 - 6)^2 + (x_2 - 5)^2 - 82.21 &\leq 0 \\ x_1 \in [13, 100], x_2 \in [0, 100] \end{aligned}$$

Table 6 shows the detailed result of the problem.

Table 6. Results of G06- Global Optimization Problem

Method	Optimal Solution	Objective Value	Constraint Viol.
Our Model	[13.0551, 0.0735]	-7883.6690	0
IPOPT	[14.0950, 0.8430]	-6961.8139	1.0×10^{-8}

4.7 Welded Beam Design

$$\text{Minimize : } f(h, l, t, b) = 1.10471h^2 + 0.04811tb(14 + l)$$

Subject to

$$\sqrt{\left(\frac{P}{\sqrt{2}hl}\right)^2 + \left(\frac{PLR}{J}\right)^2} \leq 13600$$

$$\frac{6PL}{bt^2} \leq 30000$$

$$\frac{4PL^3}{Ebt^3} \leq 0.25$$

$$P - \frac{4.013E\sqrt{\frac{bt^3}{36}}\left(1 - \frac{t}{2L}\right)}{L^2} \leq 0$$

$$h \leq b$$

Where , $R = \sqrt{4l^2 + (t + 2b)^2}$, $J = 2\sqrt{2}hl(12l^2 + (t + 2b)^2)$, $L = 14.0$, $P = 6000$, $E = 30 \times 10^6$
 $h, b \in [0.1, 2.0]$, $t, l \in [0.1, 10.0]$

The comparison between the results is given in Table 7.

Table 7. Results of Welded Beam Design Problem

Method	Optimal Solution [h, l, t, b]	Objective Value	Constraint Viol.
Our Model	[0.1696, 3.5032, 9.8944, 0.1759]	1.5769	0
IPOPT	[0.1680, 3.1905, 10.0000, 0.1680]	1.4889	1.0×10^{-4}

5. Limitations and Future Work

Incorporating the non-parametric Bayesian optimization and a parametric multi-layer perceptron (MLP) within a single methodology offers the strength associated with each paradigm, which is the proposed methodology. However, it occasionally generates solutions that are near the global optimum, which is considerable due to its appeal to penalty terms. Through errors of numerical rounding in the calculation of the gradient, the model may fail to determine the global optimum, which adds a certain level of inaccuracy and small inaccuracies may cause large deviation in objective function for certain problems. In addition to this, the method is slower than most state-of-the-art (SOTA) otherwise it may not be useful in time-sensitive scenarios but can operate efficiently with very large problems due to the ability to perform batch calculations, and many SOTA methods fail to handle problems of very large magnitude as efficiently as this method can. Future research needs to focus on improving the penalty usefulness model to lessen the effects of roundoff errors in gradient calculations. Furthermore, exploring dynamic penalty tuning and robust initialization strategies may enhance convergence to true global optima across a wider range of problems.

6. Conclusion

The suggested Bayesian self-supervised multi-layer perceptron (MLP) of a nonlinear constrained optimization is a priori feasible and adaptable in this work. Unlike in any of the conventional solvers, in the proposed methodology, a user specification for initialization is not required; instead, candidate solutions are generated from standard normal distribution and are improved further.

Such a design allows us to avoid manual intervention in the initialization and to be adaptive in different problem settings. At the core of this tool, we have Bayesian Optimization (BO) and it's not just hyperparameter tuning. It also lets the model efficiently navigate high dimension landscapes as long as exploration and exploitation are balanced right, ignoring local minima that classical solvers are easily caught in.

The model prioritizes constraint satisfaction. Inequality, equality and bound constraints are inserted directly into the loss term via differentiable penalties that serve as a stringent constraint and ensure that the optimization process is strictly feasible. Other stability improvements include gradient clipping, 2 regularization, and input-weight joint optimization to avoid oscillatory and divergent training behavior.

The experimental procedure of benchmark functionalities (Branin, Himmelblau, Rosenbrock) and engineering design problems (pressure vessel, welded beam) indicated that the proposed procedure never fails to present practical solutions, which are typically better than IPOPT since the violations are eliminated altogether. The method is not the

quickest solver, but it has the merit of being powerful in imposing feasibility and capability to avoid local traps and move towards globally competitive solutions.

Generally, the framework brings together self-supervised surrogate learning, Bayesian exploration and constraint-based optimization to the cohesion of a pipeline. The directions in the future are how to apply the method to multi-objective problems, to problems with uncertainty, and to large scale applications, in which robustness and feasibility are more important than raw speed.

References

- Alsamee, D. and Ramezani, R., Hare escape optimization algorithm with applications in engineering and deep learning, *Scientific Reports*, vol. 15, no. 1, p. 26405, 2025.
- Chen, W., Liu, S. and Tang, K., A new knowledge gradient-based method for constrained Bayesian optimization, *arXiv preprint*, 2021.
- Gu, Q., Wang, Q., Xiong, N. N., Jiang, S. and Chen, L., Surrogate-assisted evolutionary algorithm for expensive constrained multi-objective discrete optimization problems, *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 2699–2718, 2022.
- Khatamsaz, D., Vela, B., Singh, P., Johnson, D. D., Allaire, D. and Arróyave, R., Bayesian optimization with active learning of design constraints using an entropy-based approach, *npj Computational Materials*, vol. 9, no. 1, p. 49, 2023.
- Khatouri, H., Benamara, T., Breitkopf, P., Demange, J. and Feliot, P., Constrained multi-fidelity surrogate framework using Bayesian optimization with non-intrusive reduced-order basis, *Advanced Modeling and Simulation in Engineering Sciences*, vol. 7, no. 1, p. 43, 2020.
- Kilwein, Z., et al., Optimization with neural network feasibility surrogates: formulations and application to security-constrained optimal power flow, *Energies*, vol. 16, no. 16, p. 5913, 2023.
- Lei, B., Kirk, T. Q. and Bhattacharya, A., Bayesian optimization with adaptive surrogate models for automated experimental design, *npj Computational Materials*, vol. 7, no. 1, p. 194, 2021.
- Li, G., Wang, Y., Kar, S. and Jin, X., Bayesian optimization with active constraint learning for advanced manufacturing process design, *IISE Transactions*, pp. 1–15, 2025.
- Liu, X., Yang, Z., Liu, J., Xiong, J. and Huang, J., Surrogate-assisted global and distributed local collaborative optimization algorithm for expensive constrained optimization problems, *Scientific Reports*, vol. 15, no. 1, p. 1728, 2025.
- Paulson, J. A. and Tsay, C., Bayesian optimization as a flexible and efficient design framework for sustainable process systems, *Current Opinion in Green and Sustainable Chemistry*, vol. 51, p. 100983, 2025.
- Røstum, H., Gros, S. and Aas-Jakobsen, K., Constrained Bayesian optimization for engineering bridge design, *Structural and Multidisciplinary Optimization*, vol. 68, no. 1, p. 20, 2025.
- Tian, Y., Zuñiga, A., Zhang, X., Dürholt, J. P., Das, P., Chen, J., Matusik, W. and Konaković Luković, M., Boundary exploration for Bayesian optimization with unknown physical constraints (BE-CBO), *arXiv preprint*, 2024.
- Yang, Z., Jiang, C. and Liu, J., Computationally expensive constrained problems via surrogate-assisted dynamic population evolutionary optimization, *Complex & Intelligent Systems*, vol. 11, no. 2, p. 146, 2025.
- Yu, R., Picard, C. and Ahmed, F., Fast and accurate Bayesian optimization with pre-trained transformers for constrained engineering problems, *arXiv preprint*, 2024.

Biographies

Mahamudul Hassan Siddique is an undergraduate final year student of Department of Industrial and Production Engineering at BUET. He is studying the multi-tier supply chain optimization under the reinforcement learning in his thesis. His areas of research concern machine learning, deep learning, natural language processing, large language models, operations research, and optimization. He integrates the use of statistics and the AI techniques to come up with data-driven answers to complicated industrial challenges. His research is based on the interplay of conventional optimization with state-of-the-art computational models in aid of intelligent decision making.

Miftaur Rahman Zisan is an undergraduate in Industrial and Production Engineering at the Bangladesh University of Engineering and Technology (BUET). He is keenly interested in operations research, improvement in processes and computational modelling and in system analysis and optimisation techniques in addition to data-oriented solutions to intricate problems in engineering. He has experience in the application of machine learning approaches, especially with predictive analytics and decision support based upon data.

Fahimul Haque is an undergraduate final year student of Industrial and Production Engineering with expertise in statistical learning, machine learning, optimization, and computational intelligence. His research includes nonlinear and explainable machine learning models for suicide mortality prediction using high-dimensional socio-economic data, Bayesian self-supervised MLPs for initialization-free constrained optimization, and multi-agent reinforcement learning (PPO) for evolutionary game equilibrium recovery. In manufacturing systems, he develops hybrid RSM–ML frameworks, symbolic regression, and ensemble models for surface roughness and flank wear prediction in advanced drilling processes. His work emphasizes model interpretability, generalization, and decision-oriented predictive analytics.