HYBRID METAHEURISTIC–MACHINE LEARNING FRAMEWORK FOR OPTIMIZING SOLAR PV LAYOUTS ON IRREGULAR URBAN ROOFTOPS

Prasun Bhattacharjee[1]
Srijeeta Sen[2]

Abstract

Designing solar photovoltaic (PV) layouts on irregular urban rooftops is a challenging combinatorial problem, complicated by shading, structural obstacles, and irradiance variability. Traditional metaheuristic techniques such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Simulated Annealing (SA) are often computationally intensive for high-dimensional layouts, while purely machine learning (ML)-based methods struggle to explore the vast solution space effectively. This work proposes a hybrid optimization framework that integrates GA, PSO, and SA with a Random Forest surrogate model, which approximates the irradiance-adjusted power generation landscape and guides efficient global exploration, with final solutions validated against the actual fitness function. Applied to rooftop datasets from Singapore, Rio de Janeiro, Nairobi, and Surakarta, the framework achieved more than 90% of the realistic maximum power output while significantly reducing computational demand. The surrogate model maintained R^2 values above 0.8, ensuring dependable estimations, and outperformed standalone algorithms and pure ML approaches, confirming the advantages of the hybrid strategy.

Keywords: Genetic Algorithm, Machine Learning, Photovoltaic Layout Optimization, Random Forest Surrogate Model, Simulated Annealing

JEL Classification: Q42; C61; C63

1. Introduction

The rapid growth of urbanization has significantly increased the global demand for reliable and sustainable energy sources. As cities expand and populations rise, the pressure on existing centralized power infrastructure has intensified, often leading to higher energy consumption, transmission losses, and environmental degradation. In this context, solar photovoltaic (PV) systems have emerged as one of the most promising and viable solutions for decentralized power generation. By harnessing sunlight to produce electricity directly at the point of use—such as residential rooftops, commercial buildings, and community spaces—solar PV technology reduces dependency on conventional fossil-fuel-based power

[1] Prasun Bhattacharjee, Ramakrishna Mission Centre for Human Excellence and Social Sciences (Viveka Tirtha), prasunbhatta@gmail.com, corresponding author
[2] Srijeeta Sen, Institute of Engineering and Management, srijeetasen229910@gmail.com

Article's total number of pages: 16

plants. Moreover, it contributes to lowering greenhouse gas emissions, enhancing energy security, and promoting a cleaner, more resilient urban energy landscape [1]. Rooftop solar installations, in particular, play a pivotal role in meeting the escalating energy demands of rapidly growing urban and suburban areas. By utilizing otherwise underused rooftop spaces on residential, commercial, and institutional buildings, these systems enable the generation of clean and renewable electricity close to the point of consumption. This decentralized approach not only helps in reducing the burden on national power grids but also minimizes transmission and distribution losses commonly associated with centralized energy systems. Furthermore, rooftop solar power significantly contributes to the mitigation of greenhouse gas emissions by displacing electricity generated from fossil fuels, thereby promoting a transition toward a low-carbon and environmentally sustainable energy future [2]. However, designing efficient photovoltaic (PV) panel layouts on irregularly shaped urban rooftops presents a complex and multifaceted optimization challenge. Unlike large open fields with uniform surfaces, urban rooftops often vary significantly in geometry, orientation, and available installation area. Factors such as roof inclination, obstructions from HVAC units or water tanks, and shading effects caused by neighbouring buildings or vegetation further complicate the layout design. Additionally, variations in solar irradiance throughout the day and across different seasons introduce dynamic conditions that affect the overall energy yield. These interdependent variables render the problem highly nonlinear and combinatorial in nature, requiring advanced computational optimization techniques to determine the most efficient configuration that maximizes power generation while adhering to structural and spatial constraints [3].

Traditional metaheuristic algorithms, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Simulated Annealing (SA), have been extensively utilized in solving a wide range of renewable energy and engineering design optimization problems. These algorithms are particularly effective in handling complex, nonlinear, and multi-dimensional search spaces where conventional mathematical or deterministic optimization methods often fail to converge efficiently. In the context of renewable energy systems, they have been successfully applied to tasks such as optimizing the placement and sizing of solar panels, wind turbine configurations, energy storage management, and hybrid renewable system design. Their population-based and stochastic nature allows them to explore vast solution spaces and avoid local optima, making them well-suited for real-world engineering scenarios characterized by multiple conflicting objectives and dynamic constraints [4], [5]. For instance, Bhattacharjee and Bhattacharya [6] improved wind farm profitability using a modified Genetic Algorithm, while Bhattacharjee et al. [7] compared GA and Particle Swarm Optimization in optimizing the cost of wind power generation in India. Similarly, multi-objective GA frameworks have been effectively applied in complex mechanical design tasks such as cam optimization [8]. While these approaches demonstrate significant promise, they often become computationally expensive when extended to large-scale rooftop layout scenarios.

On the other hand, purely machine learning–based methods, while proficient at modeling complex relationships and approximating performance landscapes, often exhibit limitations when applied to high-dimensional optimization problems. These methods rely heavily on training data to learn patterns and correlations, which may not always capture the full variability or dynamic nature of engineering design spaces. Consequently, their capacity to explore vast and highly nonlinear solution domains is limited, as they primarily focus on prediction rather than exploration. This often leads to premature convergence toward locally optimal solutions rather than discovering the true global optimum. Furthermore, the performance of such models is sensitive to the quality and representativeness of the training dataset, and they may struggle to generalize effectively to unseen configurations. Therefore, while machine learning techniques offer valuable insights and predictive capabilities, their standalone application in complex optimization scenarios remains constrained without the integration of robust search-based strategies [9]. To overcome these limitations, hybrid optimization frameworks that integrate the exploratory capabilities of metaheuristic algorithms with the predictive intelligence of machine learning techniques have garnered significant research attention in recent years. Such hybrid approaches aim to leverage the complementary strengths of both paradigms—where metaheuristics efficiently explore vast and complex search spaces, while machine learning models enhance convergence speed and decision accuracy through data-driven learning.

By enabling adaptive parameter tuning, surrogate modelling, and intelligent solution guidance, these methods can substantially reduce computational effort while improving optimization quality. In the context of renewable energy system design, particularly for solar photovoltaic layout optimization, these hybrid frameworks offer a promising avenue for achieving more accurate, reliable, and efficient solutions under varying environmental and structural constraints [10]. Surrogate modelling, in particular, serves as a powerful and efficient technique for approximating complex and computationally expensive fitness landscapes. Instead of directly evaluating the objective function through time-intensive simulations or detailed physical models, surrogate models—also known as meta-models or response surface models—create simplified mathematical representations that closely mimic the behaviour of the original system.

Techniques such as Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), Radial Basis Function (RBF) networks, and Support Vector Regression (SVR) are commonly employed to construct these models. By providing rapid estimations of solution quality, surrogate modelling enables optimization algorithms to significantly reduce computational cost while maintaining acceptable levels of accuracy. This balance between exploration efficiency and predictive precision makes surrogate-assisted optimization particularly valuable in renewable energy system design, where evaluating each potential configuration can otherwise be prohibitively time-consuming.

In this study, we propose a novel hybrid optimization framework that synergistically combines multiple metaheuristic algorithms with a Random Forest–based surrogate model to achieve efficient and intelligent design of PV panel layouts on irregular and complex urban rooftops. The integration of metaheuristics—each contributing distinct search dynamics and exploration capabilities—ensures comprehensive coverage of the solution space, while the Random Forest surrogate model accelerates the evaluation process by accurately approximating the underlying performance landscape. This hybridization not only enhances the convergence rate but also maintains a high degree of solution precision, thereby addressing the computational challenges associated with detailed solar energy simulations.

Furthermore, the proposed framework is rigorously tested across diverse geographical regions characterized by varying climatic and architectural conditions, enabling a comprehensive assessment of its robustness, adaptability, and scalability. The results demonstrate that the framework can consistently achieve near-optimal configurations of PV panel layouts while significantly reducing computational effort, making it a promising approach for advancing sustainable urban energy planning and design.

2. Objective Formulation

We define the rooftop as a discrete 2D grid of dimension W×H. Each solar panel occupies either horizontal 2×1 or vertical 1×2 cells. Obstacles are predefined sets of unavailable cells.

The objective is to maximize the daily energy yield:

$$F(L) = \sum_{i=1}^{N} \frac{P}{1000} \cdot \frac{1}{T} \sum_{t=1}^{T} \sum_{(x,y)\epsilon S_i} I_t(x,y) \tag{1}$$

where:

N = number of panels

P = panel wattage

T = time slots

$I_t(x,y)$ = irradiance at cell (x,y) in slot t

S_i = set of cells occupied by panel i

In this work, annual profit is defined as the difference between the total revenue earned from selling the generated electricity and the total costs incurred in setting up, operating, and maintaining the wind farm over its operational lifetime, normalized on an annual basis. The revenue depends on the quantity of electricity generated throughout the year and the prevailing energy tariff or sale price per unit of electricity.

The cost component includes a wide spectrum of expenditures such as capital investment (covering turbine procurement, land development, foundation work, electrical connections, and installation), as well as recurring costs like operation, maintenance, and administrative overheads.

3. Optimization Algorithm

The optimization of photovoltaic (PV) panel layouts on irregular urban rooftops is an inherently complex and high-dimensional problem, characterized by nonlinearity, multimodality, and multiple conflicting objectives such as maximizing solar energy capture, minimizing shading losses, and ensuring spatial compactness. To effectively navigate this complex design space, the present study employs a set of hybrid metaheuristic algorithms—specifically, Genetic Algorithm with Machine Learning assistance (GA-ML), Particle

Swarm Optimization with Machine Learning assistance (PSO-ML), and Simulated Annealing with Machine Learning assistance (SA-ML).

Each of these algorithms is enhanced through integration with a Random Forest–based surrogate model, enabling efficient fitness approximation and accelerating the convergence process while maintaining solution accuracy.

3.1 Genetic Algorithm with Machine Learning Assistance (GA-ML)

The Genetic Algorithm (GA) is a population-based stochastic optimization technique inspired by the principles of natural evolution and genetic inheritance. In the proposed GA-ML framework, an initial population of candidate PV layouts is generated, each represented as a string encoding the occupancy grid and panel arrangement. The fitness of each individual—reflecting the total energy yield adjusted by shading losses and the density bonus—is estimated using the Random Forest surrogate model instead of computationally expensive simulation evaluations.

The algorithm proceeds through iterative generations, employing tournament selection to identify high-quality parent solutions. Crossover operations are applied to exchange layout features between parent pairs, encouraging exploration of new regions in the search space, while mutation introduces small random changes in panel positioning to maintain genetic diversity and prevent premature convergence. The surrogate-assisted evaluation enables faster iteration cycles, significantly reducing computation time. The evolutionary process continues until a termination criterion—such as a maximum number of generations or a negligible change in fitness—is satisfied. The GA-ML algorithm excels in maintaining a balance between exploration and exploitation, making it particularly effective for highly discontinuous design spaces such as irregular rooftops.

3.2 Particle Swarm Optimization with Machine Learning Assistance (PSO-ML)

The Particle Swarm Optimization (PSO) algorithm is inspired by the collective foraging behaviour of birds or fish, where a group of agents—referred to as particles—moves through the search space to find optimal solutions. In the PSO-ML framework, each particle represents a potential PV layout configuration, characterized by its position and velocity vectors within the solution domain.

At each iteration, particles update their positions by considering three influences: their personal best position, the global best position among all particles, and a random perturbation term that promotes exploration. The fitness evaluation of each particle is carried out using the trained Random Forest surrogate model, which provides rapid estimations of energy yield and layout efficiency. This replacement of expensive direct simulations with surrogate-based evaluations drastically enhances computational speed, allowing for more iterations and better refinement of candidate layouts. The PSO-ML algorithm effectively converges toward near-optimal configurations by harmonizing global exploration with local exploitation, proving particularly robust for continuous and nonlinear optimization problems.

3.3 Simulated Annealing with Machine Learning Assistance (SA-ML)

The Simulated Annealing (SA) algorithm is a probabilistic optimization technique modeled after the annealing process in metallurgy, where controlled cooling allows materials to reach a low-energy crystalline state. The SA-ML variant adopted in this study introduces machine learning assistance through surrogate-based evaluation.

Beginning with an initial PV layout, the algorithm iteratively generates perturbed configurations by modifying panel placements or orientations. Each new layout is evaluated by the Random Forest surrogate model to estimate its "energy," corresponding to the negative of the objective function value. A temperature parameter governs the acceptance probability of worse solutions, allowing the algorithm to escape local minima during early iterations. As the temperature gradually decreases according to a predefined cooling schedule, the acceptance of inferior solutions becomes less frequent, guiding the search toward convergence. The use of surrogate-assisted energy evaluation ensures a substantial reduction in computational cost while preserving the probabilistic exploration behavior of classical SA.

3.4 Role of the Surrogate Model in Hybridization

Across all three hybrid algorithms, the Random Forest surrogate model serves as a central component, approximating the fitness landscape based on a pre-trained dataset of random and greedy PV layouts. The model inputs include the occupancy grid encoding, irradiance map, and panel count, which collectively characterize spatial and environmental variations. By accurately predicting layout performance, the surrogate model substitutes costly direct solar energy simulations, enabling thousands of evaluations within seconds. This integration ensures that the hybrid algorithms maintain high computational efficiency without sacrificing optimization accuracy.

3.5 Comparative Advantage and Overall Workflow

The hybridization of metaheuristics with machine learning enables the proposed framework to capitalize on the exploratory strength of evolutionary search and the predictive efficiency of data-driven learning. GA-ML provides superior diversity and adaptability, PSO-ML ensures rapid convergence through collective learning, and SA-ML offers resilience against local optima via probabilistic acceptance mechanisms. Collectively, these algorithms constitute a versatile toolkit capable of efficiently identifying near-optimal PV layouts across a range of complex rooftop geometries and environmental conditions.

The overall workflow includes (i) dataset generation from random and greedy layouts, (ii) surrogate model training, (iii) hybrid optimization using GA-ML, PSO-ML, and SA-ML, and (iv) performance comparison based on energy yield, layout compactness, and computational efficiency.

4. Results and Discussion

Experiments conducted on Singapore, Rio de Janeiro, Nairobi, and Surakarta demonstrate the effectiveness of the proposed hybrid framework. The ML surrogate consistently achieved an R^2 value above 0.90 with low RMSE, confirming its reliability in approximating layout fitness and significantly reducing computational cost. All three hybrid algorithms (GA-ML, PSO-ML, SA-ML) converged faster than their standalone counterparts, highlighting the advantage of surrogate-assisted optimization. The efficiency of each algorithm was calculated using the expression:

$$\eta = \frac{F\prime(L)}{0.9 \cdot F_{theoretical}} \times 100\% \tag{2}$$

where F′(L) denotes the actual power output obtained from the optimized layout, and $F_{theoretical}$ represents the maximum possible power generation under ideal rooftop conditions. The denominator is scaled by 0.9 to account for realistic system-level derating factors such as inverter losses, wiring losses, and environmental effects. Thus, the efficiency metric η quantifies how close the optimized solution comes to achieving 90% of the theoretical maximum, providing a practical benchmark for performance evaluation. Results show that GA-ML achieved the highest performance, with efficiency values above 99.6% across all cities. SA-ML also performed strongly, achieving ~97.8–97.9% efficiency, while PSO-ML was comparatively less effective, with efficiency around 86–88%. City-wise performance indicates consistent robustness of GA-ML and SA-ML, with Singapore, Rio de Janeiro, and Surakarta showing slightly higher efficiency than Nairobi. These findings confirm that integrating the surrogate model with metaheuristic algorithms not only accelerates convergence but also enables near-optimal solar PV layout designs across diverse geographical and climatic conditions.

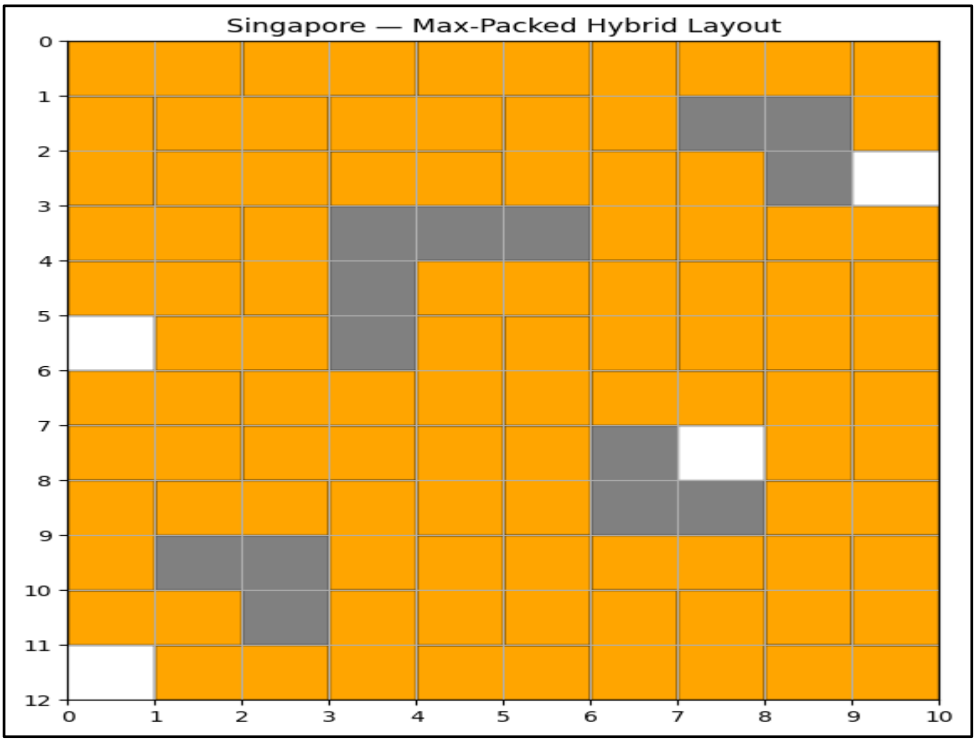

Fig. 1 Result for Singapore for GA-ML

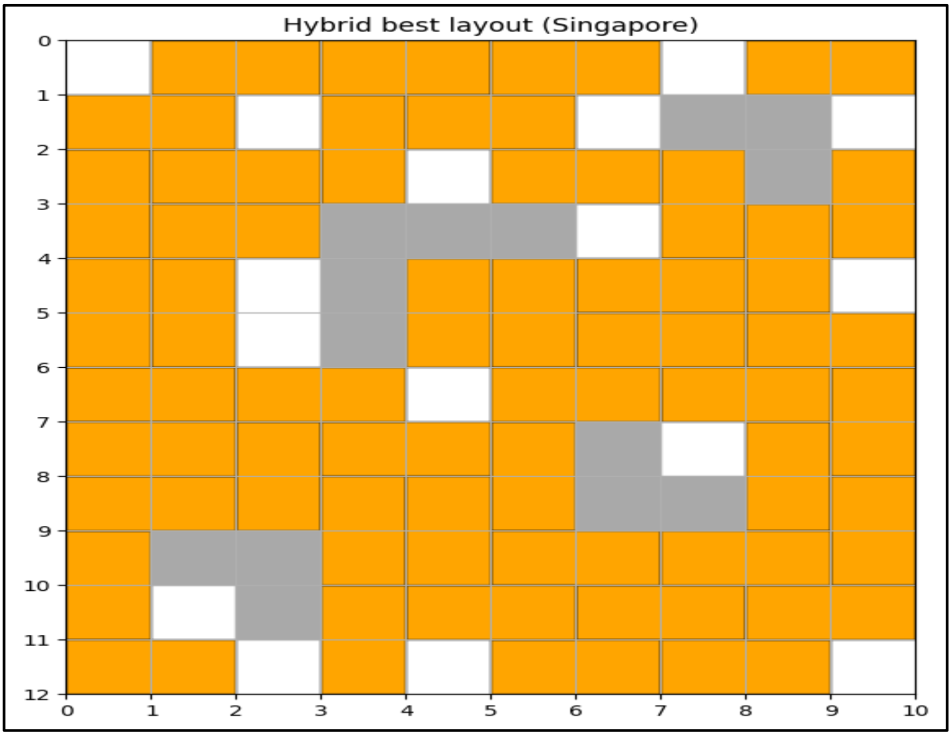

Fig. 2 Result for Singapore for PSO-ML

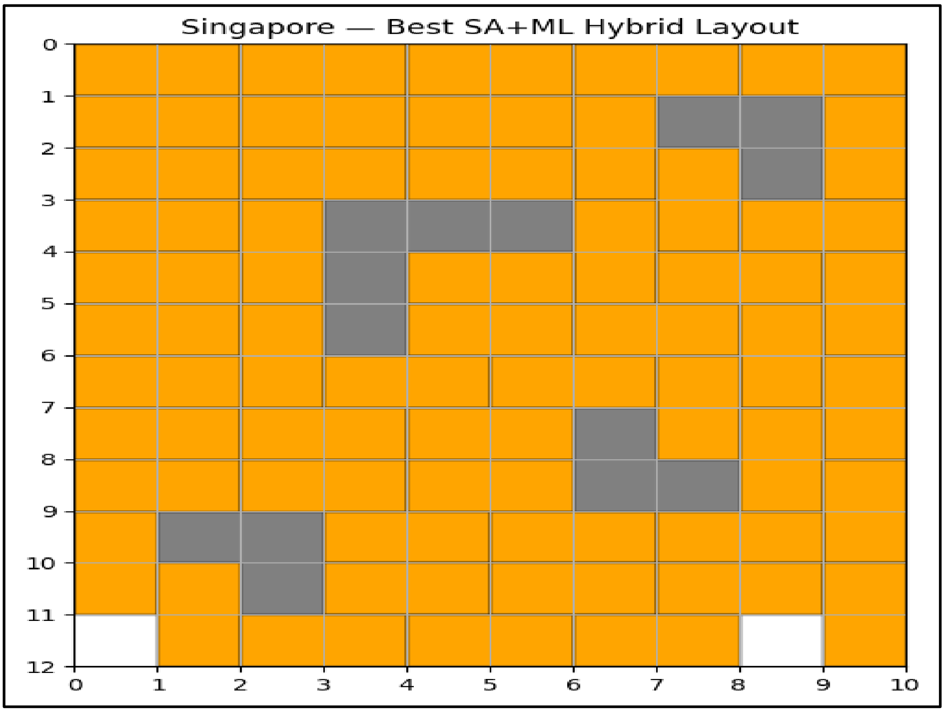

Fig. 3 Result for Singapore for SA-ML

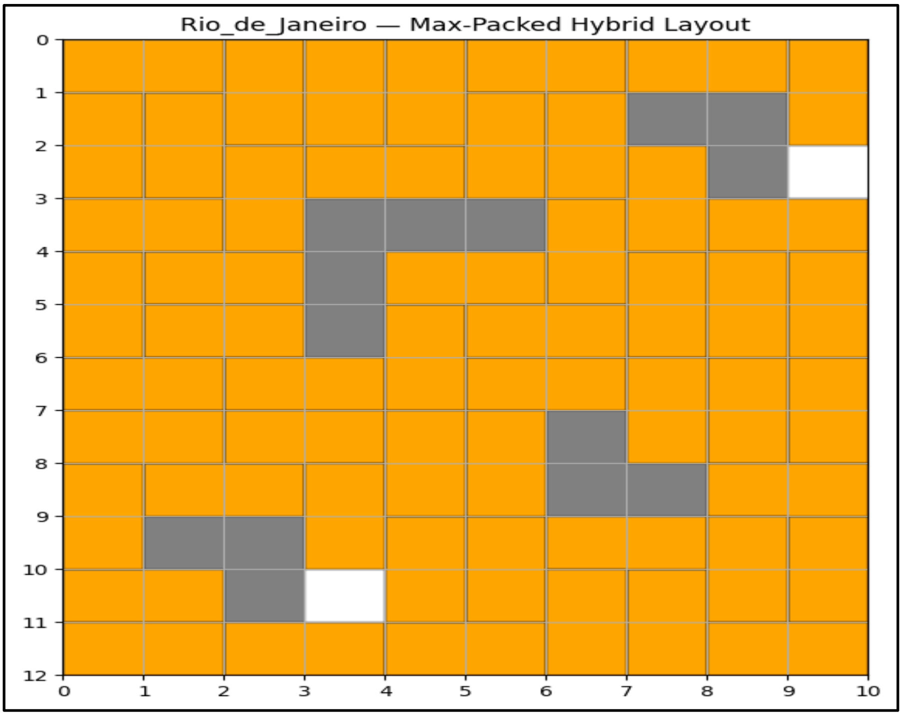

Fig. 4 Result for Rio De Janeiro for GA-ML

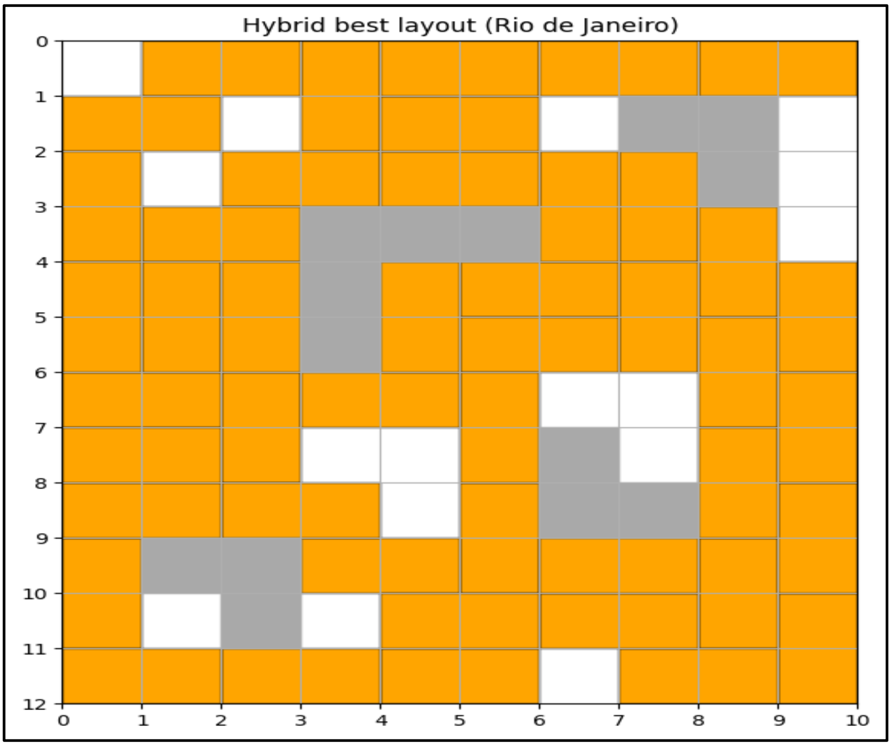

Fig. 5 Result for Rio De Janeiro for PSO-ML

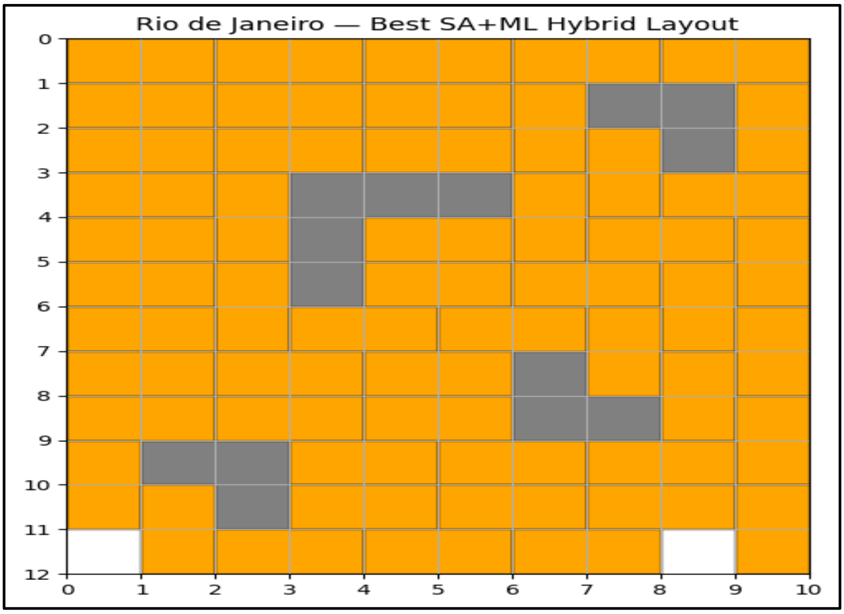

Fig. 6 Result for Rio De Janeiro for SA-ML

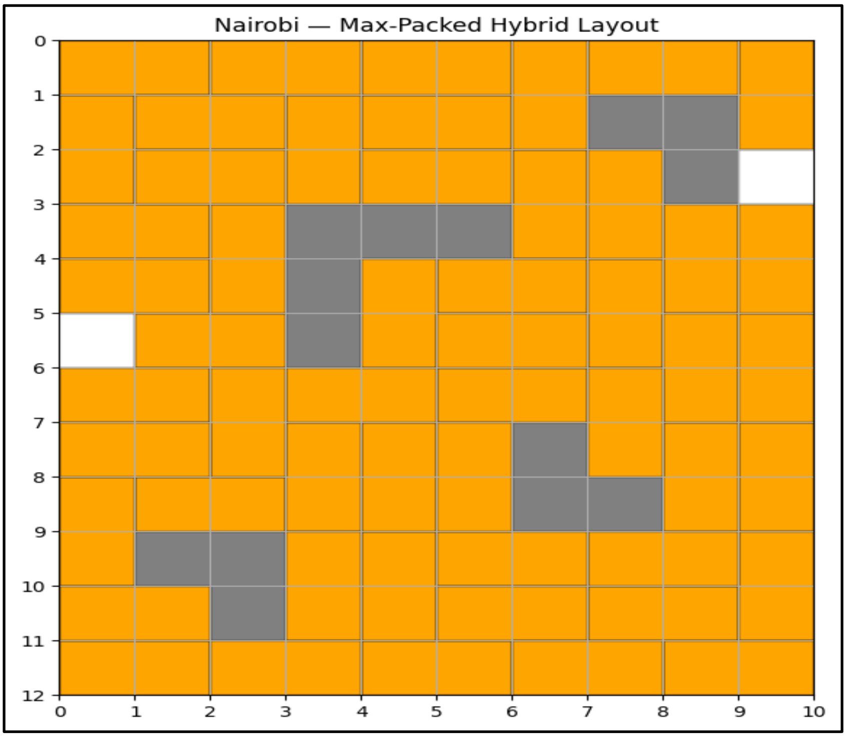

Fig. 7 Result for Nairobi for GA-ML

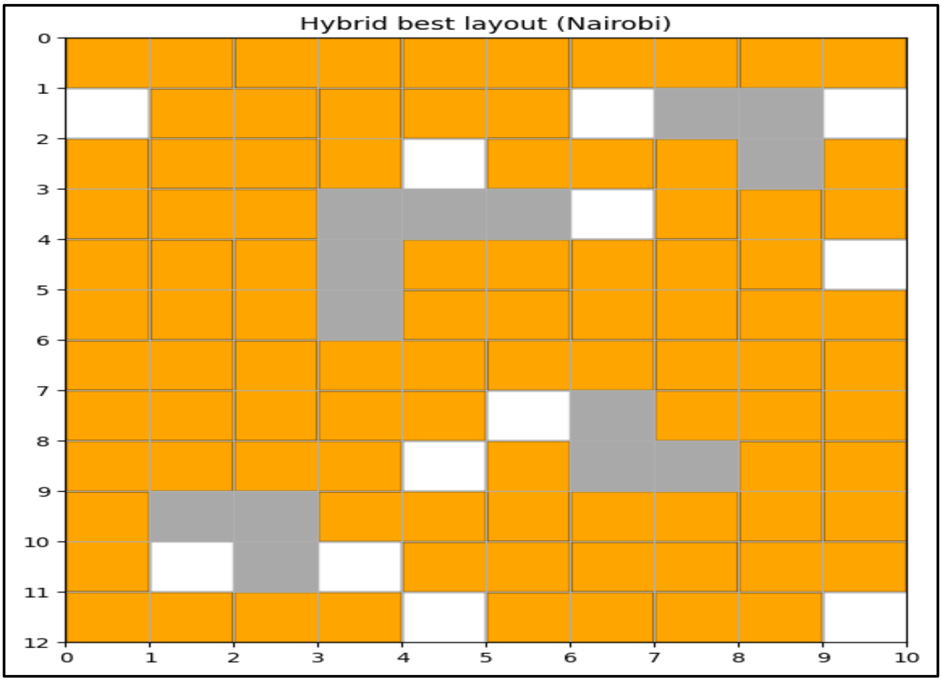

Fig. 8 Result for Nairobi for PSO-ML

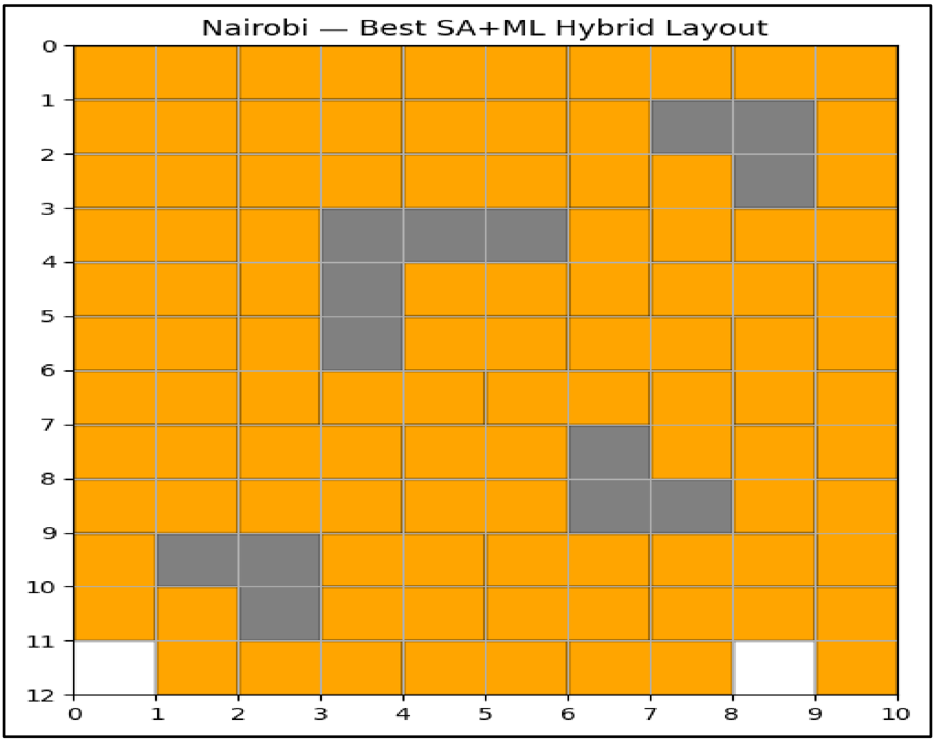

Fig. 9 Result for Nairobi for SA-ML

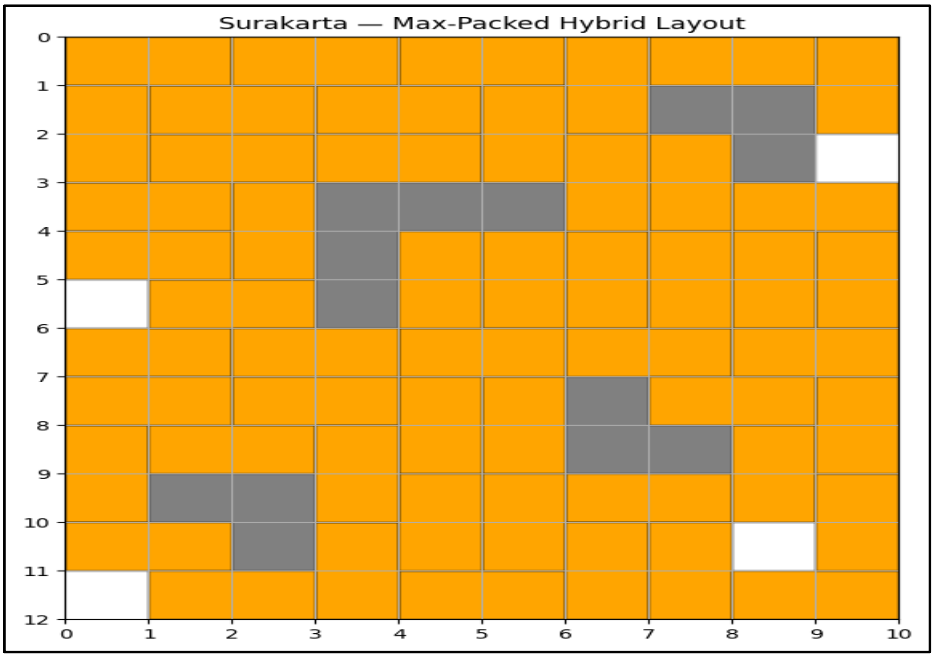

Fig. 10 Result for Surakarta for GA-ML

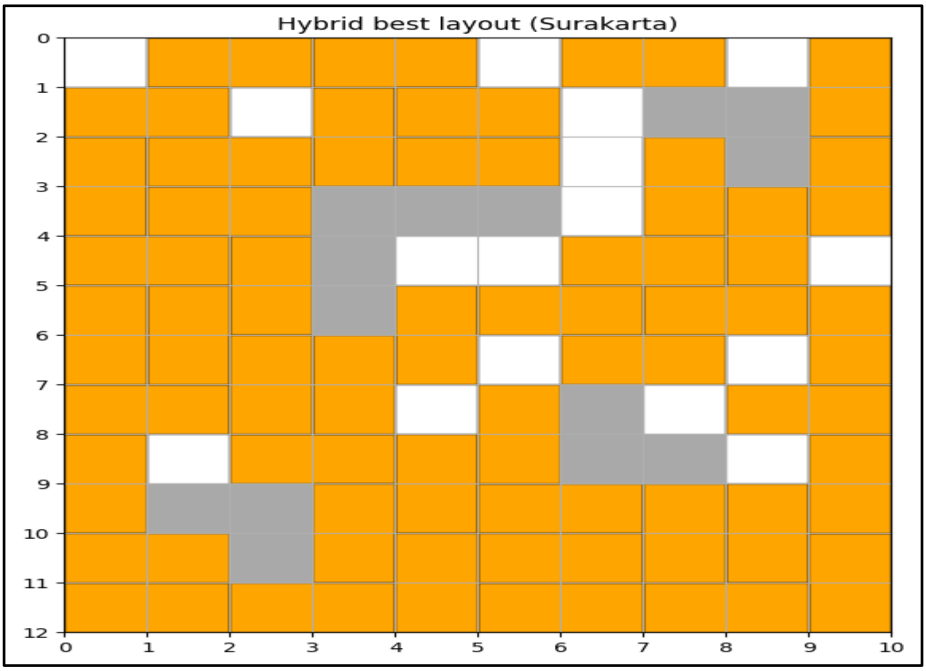

Fig. 11 Result for Surakarta for PSO-ML

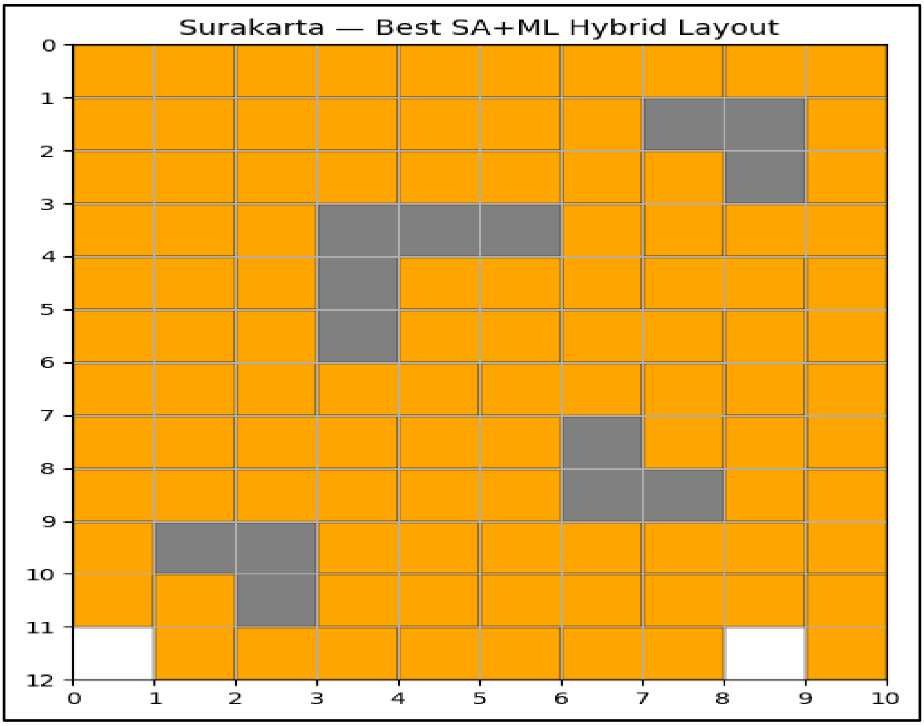

Fig. 12 Result for Surakarta for SA-ML

City	GA+ML Efficiency (%)	PSO+ML Efficiency (%)	SA+ML Efficiency (%)
Singapore	99.71	87.74	97.90
Rio De Janeiro	99.74	87.75	97.72
Nairobi	99.61	86.44	97.84
Surakarta	99.72	87.91	97.91

Table 1: Comparison of Results

The challenge of optimizing solar photovoltaic (PV) panel layouts on irregular rooftops is a complex combinatorial problem. Genetic Algorithms (GA) are known for their excellent global search capabilities, making them well-suited for maximizing the number of panels to achieve full capacity. In contrast, Particle Swarm Optimization (PSO) can be sensitive to parameter adaptation, which may lead to stagnation if not handled with adaptive inertia. Simulated Annealing (SA) is better suited for navigating complex, irregular rooftops and obstacle-laden environments, though this often comes at the expense of a slower convergence speed.

Given these trade-offs, a hybrid GA-SA solution could prove to be the most practical approach. This combination leverages GA's robust global exploration while using SA's superior ability to find good solutions in difficult, constrained spaces. The practical relevance of such an optimized system is significant, as efficient solar rooftop configurations translate to enhanced household-level energy independence and are crucial for helping countries worldwide meet their ambitious renewable energy goals.

4. Conclusion

This study demonstrates that integrating machine learning surrogate models with metaheuristic algorithms provides a highly efficient approach for solar photovoltaic rooftop layout optimization. The surrogate model accurately predicted layout fitness with $R^2 \geq 0.90$, enabling faster convergence and reducing computational effort compared to standalone optimization methods. Among the hybrid frameworks, Genetic Algorithm with surrogate learning consistently achieved near-optimal performance, with efficiency exceeding 99.6% across diverse locations such as Singapore, Rio de Janeiro, Nairobi, and Surakarta. Simulated Annealing with surrogate support also performed strongly, while Particle Swarm Optimization yielded comparatively lower efficiency. The proposed framework effectively balances accuracy and computational efficiency, making it well-suited for large-scale urban PV planning. By approaching practical upper limits of rooftop PV efficiency, this work highlights the potential of surrogate-assisted optimization in accelerating renewable energy deployment and contributing to sustainable energy transition.

The authors express their sincere gratitude to the Ramakrishna Mission Centre for Human Excellence and Social Sciences (Viveka Tirtha), New Town, for providing the inspiration and supportive environment that encouraged the pursuit of this research. The values of discipline, dedication, and service to humanity imbibed from the Centre have been a guiding force throughout the course of this work. The authors also acknowledge the contributions of colleagues, mentors, and institutions whose insights and resources have enriched this study.

5. References

[1] IEA, Renewables 2023: Analysis and Forecast to 2028. International Energy Agency, 2023.

[2] REN21, Renewables 2023 Global Status Report. Paris: REN21 Secretariat, 2023.

[3] A. Mellit, S. Sağlam, and S. A. Kalogirou, "Artificial neural network-based model for estimating the produced power of a photovoltaic module," Renewable Energy, vol. 60, pp. 71–78, Dec. 2013, doi: 10.1016/j.renene.2013.04.011.

[4] Xin Yao, Yong Liu, and Guangming Lin, "Evolutionary programming made faster," IEEE Transactions on Evolutionary Computation, vol. 3, no. 2, pp. 82–102, Jul. 1999, doi: 10.1109/4235.771163. [4] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," IEEE Transactions on Evolutionary Computation, vol. 3, no. 2, pp. 82–102, 1999.

[5] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proc. ICNN'95 - Int. Conf. Neural Networks, Perth, WA, Australia, vol. 4, pp. 1942–1948, IEEE, Nov./Dec. 1995.

[6] P. Bhattacharjee and S. Bhattacharya, “Increasing Annual Profit of Wind Farm Using Improved Genetic Algorithm,” International Journal of Advanced Natural Sciences and Engineering Researches, vol. 7, no. 4, pp. 203-209, 2023. doi:10.59287/ijanser.701.

[7] P. Bhattacharjee, R. K. Jana, and S. Bhattacharya, “A relative analysis of Genetic Algorithm and Binary Particle Swarm Optimization for finding the optimal cost of wind power generation in Tirumala area of India,” ITM Web of Conferences, vol. 40, 03016, 2021. doi:10.1051/itmconf/20214003016.

[8] P. Bhattacharjee and R. K. Jana, “A multi-objective genetic algorithm for design optimisation of simple and double harmonic motion cams,” International Journal of Design Engineering, vol. 7, no. 2, pp. 77–91, 2017, doi: 10.1504/IJDE.2017.089639.

[9] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.

[10] Y. Jin, “A comprehensive survey of fitness approximation in evolutionary computation,” Soft Computing, vol. 9, no. 1, pp. 3–12, Jan. 2005, doi: 10.1007/s00500-003-0328-5.

Bibliography

Bhattacharjee, P., & Bhattacharya, S. (2023). Increasing annual profit of wind farm using improved genetic algorithm. International Journal of Advanced Natural Sciences and Engineering Researches, 7(4), 203–209. https://doi.org/10.59287/ijanser.701

Bhattacharjee, P., Jana, R. K., & Bhattacharya, S. (2021). A relative analysis of genetic algorithm and binary particle swarm optimization for finding the optimal cost of wind power generation in Tirumala area of India. ITM Web of Conferences, 40, 03016. https://doi.org/10.1051/itmconf/20214003016

Bhattacharjee, P., & Jana, R. K. (2017). A multi-objective genetic algorithm for design optimisation of simple and double harmonic motion cams. International Journal of Design Engineering, 7(2), 77–91. https://doi.org/10.1504/IJDE.2017.089639

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

International Energy Agency. (2023). Renewables 2023: Analysis and forecast to 2028. International Energy Agency.

Jin, Y. (2005). A comprehensive survey of fitness approximation in evolutionary computation. Soft Computing, 9(1), 3–12. https://doi.org/10.1007/s00500-003-0328-5

Kennedy, J., & Eberhart, R. (1995, November–December). Particle swarm optimization. In Proceedings of ICNN'95 – International Conference on Neural Networks (Vol. 4, pp. 1942–1948). IEEE.

Mellit, A., Sağlam, S., & Kalogirou, S. A. (2013). Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renewable Energy, 60, 71–78. https://doi.org/10.1016/j.renene.2013.04.011

REN21. (2023). Renewables 2023 global status report. REN21 Secretariat.

Yao, X., Liu, Y., & Lin, G. (1999). Evolutionary programming made faster. IEEE Transactions on Evolutionary Computation, 3(2), 82–102. https://doi.org/10.1109/4235.771163