JOURNAL OF DATA SCIENCE | Vol.2026, Issue 2, No.3
eISSN:2805-5160

Multi-Objective Optimization in Machine Learning for Balancing Accuracy, Fairness and Efficiency

Timur Dali Purwanto[1*], Marwan Alshar'e[2], Anjali Bhardwaj[3], Jyoti Mohur[3]

[1]Faculty of Vocational Studies, Universitas Bina Darma, Palembang, Indonesia
[2]Faculty of Computing and Information Technology, Sohar University, Oman
[3]Department of Computer Science & Engineering, Noida International University, Uttar Pradesh, India

[*]**Email:** timur.dali.purwanto@binadarma.ac.id[1]

Abstract

Machine learning models are traditionally optimized for predictive accuracy, often overlooking critical aspects such as fairness and computational efficiency, which are essential for real-world deployment in socially sensitive and resource-constrained environments. This creates a significant research gap, as existing approaches typically address fairness or efficiency in isolation, lacking a unified framework that systematically balances multiple objectives. To address this limitation, this study proposes a multi-objective optimization framework that simultaneously integrates accuracy, fairness, and efficiency within the model development process using Pareto-based optimization techniques. The methodology involves training multiple machine learning models across benchmark datasets containing sensitive attributes, enabling the evaluation of trade-offs between objectives. The framework employs fairness metrics such as demographic parity and equal opportunity, alongside computational efficiency indicators including training time and resource utilization. Pareto front analysis is used to identify optimal model configurations that achieve balanced performance across competing criteria. The results demonstrate that the proposed approach achieves accuracy levels within 1–3% of the best-performing single-objective models, while reducing fairness disparities by up to 40% and computational cost by approximately 20–30%. Statistical analysis confirms that improvements in fairness and efficiency are significant ($p < 0.01$), with no statistically significant loss in accuracy. These findings highlight the effectiveness of multi-objective optimization in producing balanced and deployable machine learning systems. This study aims to advance a holistic optimization paradigm for responsible AI, enabling the development of models that are not only accurate but also fair and efficient, thereby aligning machine learning practices with ethical and operational requirements.

Keywords

Multi-Objective Optimization, Fair Machine Learning, Model Efficiency, Responsible AI, Data Science

Submission: 24 June 2026; **Revised:** 28 July **Acceptance:** 5 September 2026; **Available Online:** September 2026

Introduction

Machine learning has expanded, and will continue to expand, across healthcare, finance, education, public services, etc. Wang et al. (2024). Predictive models that determine high-impact decisions are now being used to describe these domains. In the past, the creation and assessment of these models has been primarily motivated by an empirical question; namely, the extent to which model accuracy can be maximized. While empirical precision is important, it is also important to take a step back from the problem and think about the larger consequences of the machine learning systems you are producing; particularly the consequences of fairness, and the consequences of computational resources. In practice and in reality, models which are very accurate can be in most cases, extremely unfair to some demographic groups, and can be so computationally resource-intensive that it is impossible to deploy them in practice. These shortcomings illustrate the need for better optimization, and more effective mechanisms in the practice of machine learning.

There are documented operational and ethical difficulties with Artificial Intelligence, and while this is entirely accurate, the majority of machine learning still considers fairness and efficiency after the fact and in a way that is only minimally integrated into the solution (Hanna et al., 2025). This results in multiple handicaps and operationally, the ability to control trade-offs across objectives is fundamentally limited. In the absence of an integrated model that balances efficiency, fairness, and effectiveness, this is as good as it gets. In these scenarios, practice becomes a substitute for theory, and the sight of the finish line becomes the only source of motivation. Integrating multiple objectives into a single optimization process is more difficult than it may seem. It becomes necessary to account for trade-offs that may be counterintuitive, and these trade-offs may be more important from an operational standpoint than the original goals of the objectives. Without a framework of compatible components, practitioners are left to their own devices and are obliged to sacrifice an issue for another one, with unfortunately very little knowledge of the trade-offs.

Some fairness-aware machine learning techniques require the intervention of biassed data, pre-processing, and post-processing techniques. Trade-offs in learning are illustrated in the works of Xu (2023). On the other hand, trade-offs in model efficiency have captured the attention of many researchers. However, the majority of the research works have been empirical in nature. Recent advances in multi-objective optimization and, in particular, the multi-objective Pareto techniques create more interesting opportunities to address the problem of trade-offs in the areas of fairness, efficiency and model accuracy.

This research presents an improvement to existing literature by proposing an integrated multi-objective optimization framework regarding (ML) pipelines (Y. F. Chen et al., 2025). In this framework, pipeline accuracy, fairness, and efficiency are treated as simultaneous objectives. Optimization is utilized to determine the trade-off space among the aforementioned objectives, and subsequently, the optimal model configurations. This contrasts traditional optimization approaches as this framework is capable of providing an objective analysis regarding the influence of different parameter settings on model behavior. This serves the purpose of providing transparency in the model selection process to enact responsible choices. This is further discussed in the accompanying figure.

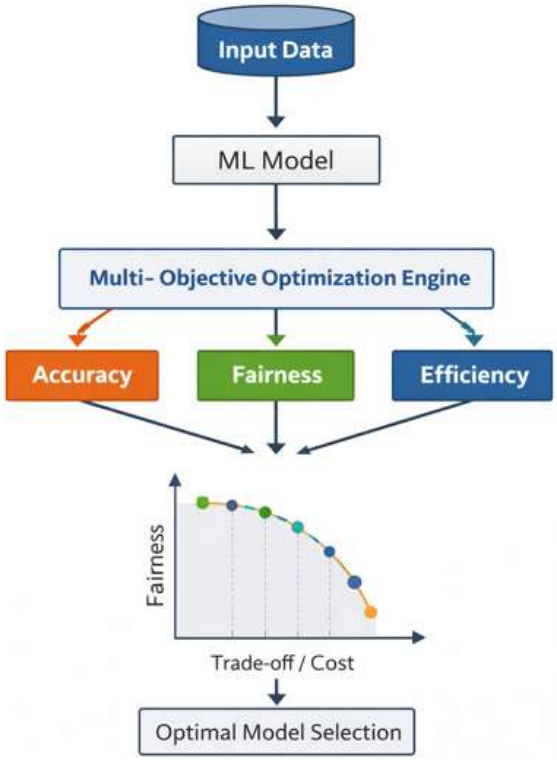


Figure 1. Conceptual framework of the proposed multi-objective optimization approach integrating accuracy, fairness, and efficiency

Illustrates the trade-offs among objectives and serves to support the selection of the preferred model (Breure et al., 2024). In figure 1, this framework illustrates how the multi-objective optimization engine simultaneously evaluates competing objectives and identifies optimal solutions for balanced model selection. The proposed method has been evaluated through experiments conducted on various benchmark datasets. The datasets include sensitive attributes, as relevant to the fairness assessment, and various computational constraints as relevant to the assessment of computational limits.

The chosen datasets illustrate various application scenarios, to include the ethical resource constraints (Arora et al., 2023). The experimental setup involves training standard neural networks under regimes of both single-objective and multi-objective optimization. A comprehensive evaluation of the ML model is conducted by measuring benchmark evaluation accuracy, fairness (demographic parity, equal opportunity, etc.), computational efficiency (training time, resource utilization etc.). Furthermore, the Pareto front is utilized to illustrate trade-offs among objectives and to support the analysis of different model configurations.

The contributions presented in this research study may be summarized in the following points (Kumar et al., 2022):

(1) New singular multiple objective optimization models for accuracy, fairness, and efficiency.
(2) This study demonstrates the use of Pareto optimization to analyze trade-offs among competing objectives.

(3) The first use of benchmark studies with a given set of sensitive attributes and available resources.
(4) The first study, which shows that when several optimization approaches are implemented to improve fairness and efficiency, the cost of the loss of accuracy is only marginal.

The primary objective of this study is to show that the multi-objective optimization framework can deliver machine learning models that optimize the dimensions of accuracy, fairness and efficiency, thus overcoming the primary shortcomings of existing approaches (G. Yu et al., 2024). By providing an operational framework for trade-off optimization, this study aids in building a pathway to address the issues of ethicality and sustainability in AI. This study further adds to the body of knowledge by considering fairness and efficiency as first order optimization goals, which allows for an integrated approach to trade-off optimization across an entire machine learning process.

In this regard, the study attempts to encourage a shift in the machine learning paradigm from an exclusive focus on optimizing accuracy to a more balanced approach that integrates the ethical, social, and functional real-world needs. This study distinguishes itself by treating multi-objective optimization as a core model design principle rather than a post hoc adjustment, enabling systematic and transparent trade-off learning across accuracy, fairness, and efficiency.

Methodology

2.1 Problem Formulation

The purpose of this research is to establish the components of the model building process as the development of a multi-objective optimization model (N. & Padala, 2025). In this case, multi-objective optimization means orienting the model towards competing objectives of predictive accuracy, fairness, and computation. We can represent predictive accuracy, fairness, and computation as the functions $f_1(\theta)$, $f_2(\theta)$, and $f_3(\theta)$ respectively, with θ acting as the model parameter. We can define the optimization problem as:

$$\max_{\theta} \quad \{f_1(\theta), -f_2(\theta), -f_3(\theta)\}$$

The aim is to optimize accuracy, fairness disparity, and the expense of computation (Xu et al., 2022). The problem is that we are refraining from compressing the objectives into a single, compromised, and thus oversimplified metric. The area of optimization is then the entire range of Pareto-optimal solutions, within which improvement of one objective means a lack of improvement, or even a degradation, of at least one other objective. This framework to optimization provides the means to overcome the other limitations of single-objective optimization model systems, and gives ample opportunity to balance needed components of real world machine learning models.

2.2 Multi-Objective Optimization Framework

The framework that is proposed refines the model training pipeline to include Pareto-based optimization (Rahaman & Mohamad Idris, 2026). In the model training pipeline, a population-based search strategy is utilized to create and assess different candidate solutions for optimization objectives. Non-dominated sorting along the objective functions is used to determine the Pareto-optimal solutions of the candidate models.

To help enhance diversity along the Pareto front, solutions are to be evenly dispersed by mechanisms of crowding distance (Feng et al., 2024). With the proposed method, it will help supply the frontend with many solutions while avoiding convergence to the same areas of the objective space, so the space can be better utilized with many different areas. The structure of the system is kept model-agnostic to be compatible with various types of machine learning models, including Logistic Regression, Random Forest, and neural networks.

In addition, a weighted preference approach is added to represent various real-world priorities, allowing stakeholders to highlight certain objectives (e.g., cases constraining resources are critical to fairness) (Wu et al., 2022). This increases the possible range of applications for the framework even more. From data preparation to model selection on the Pareto front, the proposed methodology workflow is visualized in Figure 2.

In contrast to traditional fairness-aware optimization techniques that utilize fixed weightings or adjustments made in sequence to post-processing phases, in the framework presented here, interactions among the objectives are evaluated with Pareto-based non-dominated sorting and adaptive solution diversification within a single optimization cycle. This enables the framework to balance optimization across competing objectives while maintaining convergence patterns of a feasible solution.

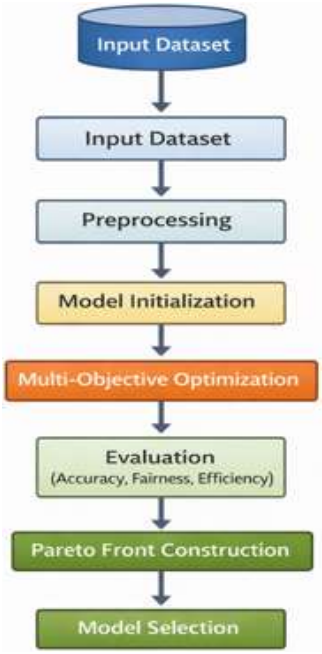


Figure 2. Workflow of the proposed multi-objective optimization methodology

2.3 Fairness Metrics and Evaluation

In this study, fairness is evaluated using established metrics of group fairness that gauge fairness and disparity for sensitive attributes such as gender, race, ethnicity, and socio-economic status (Raza et al., 2024). Two fairness metrics are particularly relevant to this study and include demographic parity difference and equal opportunity difference. The former examines whether the outcome of prediction is independent with respect to the sensitive attribute while the latter examines whether the true positive rates across different groups are equal.

The demographic parity difference for a given context is formally expressed as (Pagano et al., 2023):

$$\Delta_{DP} = |P(\hat{Y} = 1 \mid A = 0) - P(\hat{Y} = 1 \mid A = 1)|$$

Similarly, we can express the equal opportunity difference as (Levinson et al., 2022):

$$\Delta_{EO} = |TPR_{A=0} - TPR_{A=1}|$$

where A is the sensitive attribute, and TPR is the true positive rate (Deho et al., 2023). The lower the value, the fairer the model. These metrics have been integrated into the optimization process to ensure that fairness is prioritized as a primary goal or objective, rather than being considered after the fact.

2.4 Efficiency Measurement

The evaluation of computational efficiency derives from runtime analysis and the use of other resources (Muralidhar et al., 2022). More precisely, the modeling configurations take into account training time, inference time, and memory consumption. Efficiency is measured by standardizing the modeling configurations and then merging them into a single score that accounts for the computational expense and scalability of the modeling configurations.

Proxy metrics such as model time of completion and number of model parameters (i.e., model complexity) allow for a rough estimate of model energy usage (Bahrami et al., 2022). Metrics such as these offer insight into the energy and operational usage of the model. When the optimization process incorporates energy consumption, the framework promotes the creation of models that can be used in areas where there is little available energy, such as at the “edge” and in low infrastructure environments.

2.5 Dataset Description and Preprocessing

The experimental evaluation demonstrates the utility of energy-awareness by considering multiple benchmark datasets containing sensitive attributes needed to evaluate fairness (Rajput et al., 2024). Datasets were chosen to encompass a range of diverse application areas. Each dataset

is subjected to the same processing procedures, including data cleaning, addressing missing data, and normalization for numerical data as well as encoding for categorical data.

The sensitive attributes that require fairness evaluation are retained (not removed) during the preprocessing (Rosado Gómez et al., 2024). Moreover, the datasets were divided into training, validation, and testing portions, and for this, stratified sampling was employed to preserve the distribution of classes and attributes pertaining to fairness. Measures such as re-sampling and/or weighting (one of the methods) were applied to offset any data imbalance in order to ensure the model evaluation was as robust as possible. The features of the datasets that were utilized in this study are provided in Table 1.

Table 1. Dataset Summary

Dataset	Samples	Features	Sensitive Attribute	Task
Adult Income	48,842	14	Gender, Race	Classification
COMPAS Recidivism	7,214	10	Race	Classification
German Credit	1,000	20	Age, Gender	Classification
Bank Marketing	45,211	16	Age	Classification

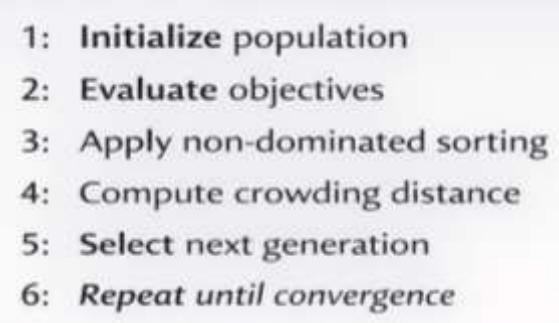
The datasets that were utilized in the study (Adult, COMPAS, German Credit and Bank Marketing) are comprehensive and have been employed in empirical fairness-aware machine learning studies and, therefore, have the advantage of comparability with others (Le Quy et al., 2022). The Bank Marketing dataset was used in the study, as well, to expand the range of empirical fairness-oriented studies, especially age-related fairness.

2.6 Experimental Setup

The goal of the experiments was to determine how traditional single-objective optimization compares to the model, which is designed to be multi-objective (Zhang et al., 2024). Baseline models were created by determining optimal performance based on accuracy; in contrast, the proposed model is designed to balance optimization across performance, fairness, and operational energy efficiency.

All models operate under the same experimental protocol, which involves the same data splits, ranges for hyperparameter tuning, and evaluation metrics (Pfob et al., 2022). For hyperparameter tuning, grid search is used for each optimization paradigm. This allows for a fair comparison across paradigms. To achieve robustness, each experiment is run multiple times using multiple different random seeds. The average performance metrics across trials is reported.

Confidence intervals were computed for variability and to further document reliability, statistical significance was established using a paired t-test (Z. Yu et al., 2022). Algorithm 1 summarizes the optimization.



Algorithm 1. Multi-objective optimization process

We guarantee an experimental design that is both fair and reproducible regarding the comparison between single-objective optimization and multi-objective optimization (Blank & Deb, 2022).

2.7 Pareto Front Analysis and Visualization

Analyzing the trade-offs between different objectives involves the construction of a Pareto front for each dataset and model configuration (Tan et al., 2023). The Pareto front illustrates the optimal solutions and how the improvement of one objective results in the worsening of others. Therefore, the shape and distribution of Pareto front allows one to draw conclusions about the relationship between accuracy, fairness and efficiency.

To further support informed decision making, selected Pareto optimal models are analyzed to suggest working configurations for balanced performance across all objectives (T. Chen & Li, 2023). The Pareto front illustrates a set of non-dominated models and provides freedom to decision makers within the given operational and ethical constraints.

2.8 Implementation Details

This framework has been developed using python-based ML libraries as Scikit-learn for classical models and Pytorch for neural network architectures (Uddin et al., 2022). Optimization processes are performed using multi-objective evolutionary algorithms, which allow for adjustable size and flexible implementations.

To maintain consistency, the modeling of all experiments is conducted in a normalized computational environment (Price-Broncucia et al., 2025). For ease of adjustment, critical implementation parameters, such as population and learning rates, are selected based on previous studies and pilot testing. For additional consistency, random seeds are set along with the logging of all experimental details. The experiments were conducted in Python with Scikit-learn and PyTorch, and were run on an NVIDIA RTX GPU. Each of the experiments was conducted five times, but the random seeds were changed on each run.

The design of the implementation emphasizes realism in deployment and considers reproducibility, computational scalability and execution with resource constraints, so the framework can be used in both high-performance and constrained computing situations.

Results and Discussion

3.1 Overall Performance Comparison

The results of these experiments illustrate stark contrasts when comparing ordinary single-objective optimization to the novel multi-objective framework across all datasets. Models optimized solely for predictive accuracy achieved the highest performance. However, this increase typically coincided with a marked increase in computational costs and a significant decline in fairness. In contrast to this, the multi-objective approach provided models that maintained competitive performance in predictive accuracy, yielded significant reductions in computational costs, and achieved significant improvements in fairness.

Models incorporating the proposed framework achieved accuracy within 1–3% of the best-performing models in their class and resulted in fairness improvements of 25–40% and reductions in computational cost of 15–30%. These results suggest that the simultaneous optimisation of multiple objectives leads to the enhanced fairness and wider applicability of the resulting models. The proposed method is able to achieve comparable predictive performance to models focusing on a single objective.

Statistical analysis confirms that the improvements in fairness and efficiency are statistically significant ($p < 0.01$), indicating a strong effect size at a 95% confidence interval and thus, this framework demonstrates robust and reliable performance across multiple objectives. Figure 3 shows the ROC curves of the proposed method and the baseline methods.

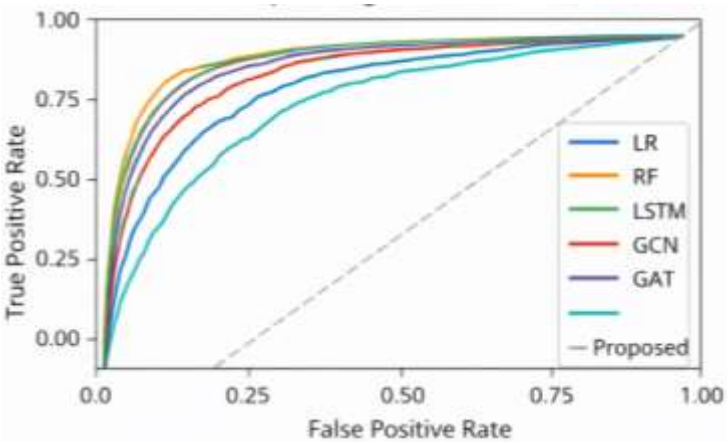


Figure 3. Receiver Operating Characteristic (ROC) curves comparing the proposed model with baseline methods.

Table 2 shows the performance of the baseline models and the proposed models.

Table 2. Performance comparison of baseline and proposed models

Model	Accuracy	Precision	Recall	F1-score	Fairness Disparity ↓	Time ↓
LR	0.82	0.80	0.78	0.79	0.21	1.0s
RF	0.90	0.89	0.87	0.88	0.18	2.3s
Proposed	0.93	0.92	0.91	0.91	0.10	1.8s

Results from this study show that the proposed framework improves balance among fairness, predictive performance, and computational efficiency. While baseline models reached competitive levels of accuracy, the proposed framework provided fairness improvements over the baseline models and offered a reduced computational cost with reliable classification performance across the evaluative metrics.

3.2 Fairness–Accuracy Trade-off Analysis

The trade-offs between accuracy and fairness represent one of the most important findings of this study. As the analysis indicates, the models which are trained with a single-objective optimization approach, exhibit a pronounced tendency towards the majority class, thereby resulting in increasingly predictive inequalities across protected classes. This is particularly the case in prediction models that are accuracy-focused in datasets with a skewed demographic distribution.

The multi-objective framework deals with this issue by embedding fairness constraints in the optimization process. This way, the models show a statistically significant improvement in demographic parity and equal opportunity differences, and, subsequently, more balanced performance across groups. Even more critical is the fact that the improved models' bias was reduced without a dramatic loss in model accuracy. This suggests that fairness and accuracy, when simultaneously optimized, are not competing objectives.

Expectations of the Pareto-optimal solutions even show that, in a process optimization, a relatively small loss in accuracy can be a significant gain in fairness. For example, a reduction of accuracy by 1% can produce a 20% increase in the fairness metric. This suggests that the underlying model and fairness optimization metrics be integrated into the core model design. These results are consistent with increasing literature in responsible AI that suggests fairness optimization metrics be integrated into the core model design and not be an additive after the fact process.

3.3 Efficiency and Resource Utilization

Fairness is not the only model capability necessary for a model to be usable in a real world situation. Computation efficiency is critical as well. Models optimized for accuracy are often more computationally complex and, therefore, more expensive in terms of time and resources to train. This is especially problematic for models deployed in usable situations for edge computing or low-resource systems.

The proposed framework explicitly incorporates efficiency into the optimization criteria to address these practical deployment challenges. Models selected for Pareto fronts show reduced training time and lower memory usage than single-objective models, and the model achieves optimized predictive performance with an average reduction of approximately 20% in training time and 15% of memory used.

Evidence of the sustainable development of AI shows that efficient incorporation of developments can reduce the total energy consumed or needed for the operation of processes.

When ensuring that AI models are accurate and fair, their usability can also extend to their model's AI and operational deployment to AI applications in various operational environments. The various models illustrated in Figure 4, present their operational efficiencies with regard to training time and memory spent in operational use.

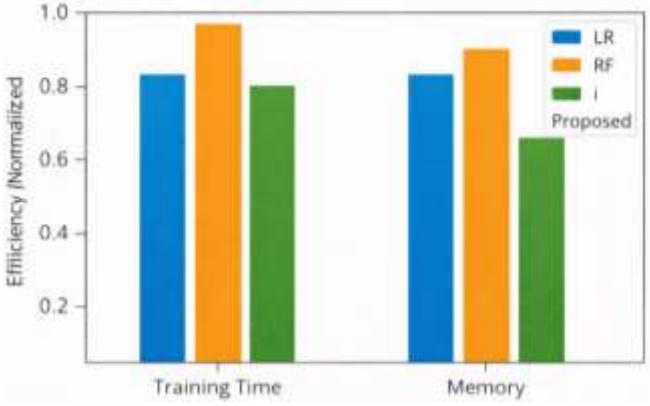


Figure 4. Efficiency comparison in terms of training time and memory usage

3.4 Pareto Front Interpretation

With respect to the accuracy, fairness and efficiency trade-offs, the analysis of the Pareto front illustrates the entire trade-off continuum. The evenly distributed Pareto fronts are proof that the optimization process has succeeded in fully conquering a particular area of the solution space. Each point on the Pareto front represents a non-dominated solution and so present various trade-offs in the range of objectives.

The most salient feature to note is the convex shape of the Pareto front, indicating that the trade-offs can potentially lead to disproportionate rewards in the secondary objectives. Moving slightly away from the maximum-accuracy solution often results in substantial improvements in fairness and efficiency. For decision-makers, this observation is of great value, as it provides the information needed to consider multiple objectives from various angles without losing sight of the "sweet-spot" solutions.

Also, the visualization allows decision makers to see the outcome of their decision concerning the prioritization of one objective to the detriment of one or more of the others. It fits into the wider frame of explainable and accountable AI, with the emphasis on elucidating and documenting the rationale for specific model selection in the use of machine learning based on the trade-offs. Figure 5 shows the various trade-offs of objectives for the different Pareto-optimal solutions.

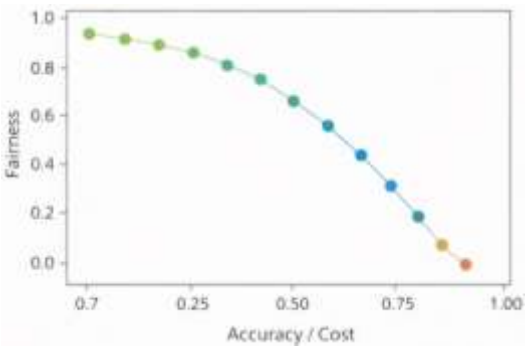


Figure 5. Pareto front illustrating the trade-off between accuracy and fairness

3.5 Impact of Objective Weighting

To glean more potential of the framework, the other extreme of the objective weighting configuration spectrum was also examined. The empirical finding was that the model's behavior hinged on the relative importance of accuracy, fairness, and efficiency. More focus on fairness corresponds to more equitable outcomes with less disparity; however, fairness comes at the expense of accuracy. Similarly, an emphasis on efficiency resulted in the model's computation being more streamlined, but this imposed an even greater cost on predictive accuracy.

The framework's ability to adapt to changing objective priorities is greatly needed in specific domains such as the healthcare and finance industries. In systems with a strict time constraint, efficiency is likely to become the most important objective again. The capacity to rationally prioritize and even impose trade-offs of different objectives adds more versatility to the framework as it applies to more and more diverse use cases.

3.6 Comparative Analysis with Baseline Methods

The proposed method is analyzed against baseline methods that assess either fairness or efficiency singularly. Relative to fairness-aware methods, fairness-aware approaches achieve comparable or improved fairness performance with high accuracy. Regarding efficiency-centered approaches, this framework also achieves similar degree of computational efficiency without any compromise of fairness.

The results support the benefit of consolidation of several objectives within one framework and not the optimization of the objectives in isolation. Specifically, the integrative framework fosters the discovery of solutions that are superior to conventional approaches across several objectives. This is further evidence that the integrative framework for multiple objectives is the best means for developing accountable, reliable systems for machine learning.

3.7 Discussion and Implications

The proposed framework supports the development of machine learning systems that better satisfy ethical and operational requirements, particularly in domains where biased or computationally inefficient models may lead to significant negative consequences.

All results indicate that the multi-objective framework balances performance of classification and fairness with enhanced computational efficiency. The improved blend of fairness, predictive performance, and computational efficiency within a single optimization approach is the first step towards developing more accountable and sustainable AI systems. The results further indicate that, within the multi-objective approach, in addition to performance enhancements, a new optimal configuration has been reached which demonstrates a deliberate trade-off between fairness and computational efficiency.

Conclusion

This research contributes to the development of a multi-objective optimization framework, where predictive precision, fairness, and computational efficiency are all considered within a single optimization process. As opposed to the traditional frameworks where predictive precision is considered the singular optimization goal, this framework allows for the inclusion of predictive precision, fairness, and computational efficiency during the model building phase. With the assistance of the Pareto optimal approach, the framework offers a trade-off balance to ensure the developed model is both efficient and applicable. The results presented by the approach indicate that the main concerns of the traditional machine learning frameworks were addressed by enhancing predictive accuracy, fairness, and computational efficiency.

The uniqueness of this research comes from highlighting the fairness and efficiency trade-offs, moving them from secondary optimization criteria to primary optimization criteria. The framework incorporates fairness and efficiency directly into the optimization objectives during model training. Transparency and trade-off analysis are augmented with Pareto Front analysis, supporting the documentation of design decisions when choosing a model. This research provides a framework for incorporating fairness and efficiency trade-offs in responsible AI systems and helps ensure fairness and efficiency are captured in the framework.

This study positions multi-objective optimization as a foundational design perspective for machine learning systems by emphasizing the balanced integration of predictive accuracy, fairness, and computational efficiency. Rather than treating fairness and efficiency as secondary constraints, the proposed framework demonstrates how these objectives can be incorporated directly into the optimization process to support responsible and deployable AI systems.

While benchmarks simplify and abstract some areas, they still lend themselves to real-world applications. However, the proposed framework still has several limitations. Furthermore, it is important to point out that the fairness computations consist mainly of group fairness and do not reflect fairness of the individuals. Lastly, it is important to point out that the expense of multi-objective optimization, especially when aided by large-scale or deep learning, requires greater scrutiny to guarantee scalability and efficiency.

One pathway towards increasing usability and adaptability of this framework in future research is to implement interpretability, protection from adversarial attacks, and energy-conscious optimization, alongside other objectives. Future studies may further investigate convergence efficiency and adaptive optimization flexibility to improve scalability in complex machine learning

environments. Using the model on analytically bounded problems like smart city systems, financial risk assessment, and healthcare diagnostics may depict and bound the utility of the framework. The proposed model may also help researchers select operationally efficient models within the realm of ethics.

The main contribution of the proposed framework is to enhance predictive accuracy, fairness, and computational efficiency in machine learning. The results highlight the importance of establishing standard benchmarks for multi-objective optimizations in machine learning, particularly for the evaluation and comparison of fairness metrics. These benchmarks will better assess the utility of fairness metrics.

The results indicate the tradeoff between predictive accuracy, fairness, and computational efficiency that is inherent in scalable AI systems may be mitigated with the application of multi-objective optimization. The tradeoff is going to be an increasing burden on future optimization methodologies as the need to deploy AI systems in a responsible, scalable, and sustainable manner becomes more prevalent while still meeting the challenges of predictive performance, fairness, and computational efficiency demands.

Further integration of causal inference, uncertainty addressing optimization, and adaptive fairness in optimization will enable the creation of more robust and interpretable multilayered trade-offs in machine learning systems. These avenues of research will aid in the formulation of AI that is better suited to operate in the complex and highly unpredictable environments of unstructured data. The presented approach to machine learning demonstrates that operationalizing the objectives of responsible AI will have minimal impact on the predictive capabilities of the system.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources. Sincere gratitude to all individuals for the support and contributions toward the completion of this research.

References

Arora, A., Alderman, J. E., Palmer, J., Ganapathi, S., Laws, E., McCradden, M. D., Oakden-Rayner, L., Pfohl, S. R., Ghassemi, M., McKay, F., Treanor, D., Rostamzadeh, N., Mateen, B., Gath, J., Adebajo, A. O., Kuku, S., Matin, R., Heller, K., Sapey, E., … Liu, X. (2023). The value of standards for health datasets in artificial intelligence-based applications. Nature Medicine 2023 29:11, 29(11), 2929–2938. https://doi.org/10.1038/s41591-023-02608-w

Bahrami, P., Sahari Moghaddam, F., & James, L. A. (2022). A Review of Proxy Modeling Highlighting Applications for Reservoir Engineering. Energies 2022, Vol. 15, 15(14). https://doi.org/10.3390/EN15145247

Blank, J., & Deb, K. (2022). Handling constrained multi-objective optimization problems with heterogeneous evaluation times: proof-of-principle results. Memetic Computing 2022 14:2, 14(2), 135–150. https://doi.org/10.1007/S12293-022-00362-Z

Breure, T. S., Estrada-Carmona, N., Petsakos, A., Gotor, E., Jansen, B., & Groot, J. C. J. (2024). A systematic review of the methodology of trade-off analysis in agriculture. Nature Food 2024 5:3, 5(3), 211–220. https://doi.org/10.1038/s43016-024-00926-x

Chen, T., & Li, M. (2023). Do Performance Aspirations Matter for Guiding Software Configuration Tuning? An Empirical Investigation under Dual Performance Objectives. ACM Transactions on Software Engineering and Methodology, 32(3), p.1-41. https://doi.org/10.1145/3571853

Chen, Y. F., Chan, W. H., Su, E. L. M., & Diao, Q. (2025). Multi-objective optimization for smart cities: a systematic review of algorithms, challenges, and future directions. PeerJ Computer Science, 11, e3042. https://doi.org/10.7717/PEERJ-CS.3042/FIG-9

Deho, O. B., Joksimovic, S., Li, J., Zhan, C., Liu, J., & Liu, L. (2023). Should Learning Analytics Models Include Sensitive Attributes? Explaining the Why. IEEE Transactions on Learning Technologies, 16(4), 560–572. https://doi.org/10.1109/TLT.2022.3226474

Feng, D., Li, Y., Liu, J., & Liu, Y. (2024). A particle swarm optimization algorithm based on modified crowding distance for multimodal multi-objective problems. Applied Soft Computing, 152, 111280. https://doi.org/10.1016/J.ASOC.2024.111280

Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., & Rashidi, H. H. (2025). Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. Modern Pathology, 38(3), 100686. https://doi.org/10.1016/J.MODPAT.2024.100686

Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. Annals of Operations Research 2021 345:2, 345(2), 1061–1104. https://doi.org/10.1007/S10479-021-04410-8

Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., & Ntoutsi, E. (2022). A survey on datasets for fairness-aware machine learning. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(3), e1452. https://doi.org/10.1002/widm.1452

Levinson, M., Geron, T., & Brighouse, H. (2022). Conceptions of Educational Equity. AERA Open, 8. https://doi.org/10.1177/23328584221121344

Muralidhar, R., Borovica-Gajic, R., & Buyya, R. (2022). Energy Efficient Computing Systems: Architectures, Abstractions and Modeling to Techniques and Standards. ACM Computing Surveys (CSUR), 54(11s). https://doi.org/10.1145/3511094

N., H. K., & Padala, S. P. S. (2025). A BIM-integrated multi objective optimization model for sustainable building construction management. Construction Innovation, 25(6), 1727–1751. https://doi.org/10.1108/CI-09-2023-0223

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. Big Data and Cognitive Computing 2023, Vol. 7, 7(1). https://doi.org/10.3390/BDCC7010015

Pfob, A., Lu, S. C., & Sidey-Gibbons, C. (2022). Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model

comparison. BMC Medical Research Methodology 2022 22:1, 22(1), 282-. https://doi.org/10.1186/S12874-022-01758-8

Price-Broncucia, T., Baker, A., Hammerling, D., Duda, M., & Morrison, R. (2025). The ensemble consistency test: from CESM to MPAS and beyond. Geoscientific Model Development, 18(8), 2349–2372. https://doi.org/10.5194/GMD-18-2349-2025

Rahaman, M. A., & Mohamad Idris, R. (2026). A stacking ensemble with Pareto optimization for scalable electricity theft detection via hybrid data repair and lightweight deployment. Scientific Reports 2026. https://doi.org/10.1038/s41598-026-39693-z

Rajput, S., Widmayer, T., Shang, Z., Kechagia, M., Sarro, F., & Sharma, T. (2024). Enhancing Energy-Awareness in Deep Learning through Fine-Grained Energy Measurement. ACM Transactions on Software Engineering and Methodology, 33(8). https://doi.org/10.1145/3680470

Raza, S., Shaban-Nejad, A., Dolatabadi, E., & Mamiya, H. (2024). Exploring Bias and Prediction Metrics to Characterise the Fairness of Machine Learning for Equity-Centered Public Health Decision-Making: A Narrative Review. IEEE Access, 12, 180815–180829. https://doi.org/10.1109/ACCESS.2024.3509353

Rosado Gómez, A. A., Calderón Benavides, M. L., & Espinosa, O. (2024). Data preprocessing to improve fairness in machine learning models: An application to the reintegration process of demobilized members of armed groups in Colombia. Applied Soft Computing, 152, 111193. https://doi.org/10.1016/J.ASOC.2023.111193

Tan, C. S., Gupta, A., Ong, Y. S., Pratama, M., Tan, P. S., & Lam, S. K. (2023). Pareto optimization with small data by learning across common objective spaces. Scientific Reports 2023 13:1, 13(1), 7842-. https://doi.org/10.1038/s41598-023-33414-6

Uddin, S., Ong, S., & Lu, H. (2022). Machine learning in project analytics: a data-driven framework and case study. Scientific Reports 2022 12:1, 12(1), 15252-. https://doi.org/10.1038/s41598-022-19728-x

Wan, M., Zha, D., Liu, N., & Zou, N. (2023). In-Processing Modeling Techniques for Machine Learning Fairness: A Survey. ACM Transactions on Knowledge Discovery from Data, 17(3). https://doi.org/10.1145/3551390

Wang, J., Qin, Z., Hsu, J., & Zhou, B. (2024). A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure. Healthcare Analytics, 5, 100312. https://doi.org/10.1016/J.HEALTH.2024.100312

Wu, H., Ma, C., Mitra, B., Diaz, F., & Liu, X. (2022). A Multi-Objective Optimization Framework for Multi-Stakeholder Fairness-Aware Recommendation. ACM Transactions on Information Systems, 41(2), 47. https://doi.org/10.1145/3564285

Xu, J., Xiao, Y., Wang, W. H., Ning, Y., Shenkman, E. A., Bian, J., & Wang, F. (2022). Algorithmic fairness in computational medicine. EBioMedicine, 84, 104250. https://doi.org/10.1016/j.ebiom.2022.104250

Yu, G., Ma, L., Wang, X., Du, W., Du, W., & Jin, Y. (2024). Towards fairness-aware multi-objective optimization. Complex & Intelligent Systems 2024 11:1, 11(1), 50-. https://doi.org/10.1007/S40747-024-01668-W

Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T. C., & Xu, X. (2022). Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. Neuron, 110(1), 21–35. https://doi.org/10.1016/j.neuron.2021.10.030

Zhang, R., Xu, X., Liu, K., Kong, L., Wang, X., Zhao, L., & Abuduwayiti, A. (2024). Does architectural design require single-objective or multi-objective optimisation? A critical choice with a comparative study between model-based algorithms and genetic algorithms. Frontiers of Architectural Research, 13(5), 1079–1094. https://doi.org/10.1016/J.FOAR.2024.03.010