Enhancing the Prediction of Multiple Ozone Metrics Using Genetic Algorithm-Based Feature Selection for the Multi-Target Regression of the Environmental AQ-Bench Dataset

N. Mohamed Abdul Kader Jailani
School of Computer Science and Applications, REVA University, Bangalore, India
jailani.msa@gmail.com (corresponding author)

Geeta C. Mara
School of Computing and Information Technology, REVA University, Bangalore, India
geetac.mara@reva.edu.in

Received: 21 September 2025 | Revised: 29 October 2025 | Accepted: 6 November 2025

DOI: https://doi.org/10.48084/etasr.14985

ABSTRACT

Predicting multiple air quality metrics from high-dimensional environmental datasets is a significant challenge hampered by the "curse of dimensionality". The present study introduces and evaluates three novel Genetic Algorithm (GA)-based feature selection methodologies designed specifically for Multi-Target Regression (MTR) tasks using the AQ-Bench dataset. The proposed wrapper-based approach integrates GAs with MTR models to identify optimal feature subsets. The results demonstrate a substantial reduction in feature dimensionality, by up to 61.6%, while concurrently improving predictive performance over baseline models. This research establishes a practical framework for practitioners, showing that a common feature subset (GA-FS-MTR) is effective for correlated targets, whereas a per-target approach (GA-FS-TARGET) excels when precision for heterogeneous targets are required. A key finding of the present study is the identification of a structural sensitivity in complex models like the Ensemble of Regressor Chains (ERC), where global optimization can inadvertently remove features vital for its chained architecture. This work validates GA-based feature selection as an effective tool for optimizing MTR models in environmental science and provides a strategic guide for its implementation.

Keywords-feature selection; genetic algorithm; Multi-Target Regression (MTR); air quality prediction; Single Target (ST); Stacked Single Target (SST); Ensemble of Regressor Chains (ERC); Average RRMSE

I. INTRODUCTION

In the contemporary era of big data, data-driven decision-making has become a cornerstone of progress across nearly every scientific and industrial domain. Machine learning, and specifically regression analysis, provides the fundamental tools for modeling relationships from complex datasets [1]. While traditional regression focuses on a single output, an increasing number of real-world problems require the simultaneous prediction of multiple, often correlated, target variables. This problem is addressed by MTR, a powerful machine learning paradigm with broad applicability in domains such as environmental science [2], healthcare [3], telecommunications [1], agriculture [4], neuroscience [5, 18], epidemiology [7], and the mapping of simultaneous activities in different brain regions for neurological studies [6].

The primary advantage of MTR models is their ability to leverage inter-target dependencies, potentially yielding more accurate and holistic predictions than separate single-target models [8]. However, this capability introduces significant challenges, which are severely exacerbated when dealing with high-dimensional input data [5]. As the number of input features grows, data become increasingly sparse, leading to the "curse of dimensionality" [5]. This phenomenon results in several adverse outcomes: models become prone to overfitting [9], computational costs increase dramatically, and, critically, interpretability diminishes [10]. A model with hundreds of features becomes a "black box," making it difficult for domain experts to trust its predictions or derive insights [8, 9,11, 12].

Despite the impact of feature quality on performance [13], most research on MTR has prioritized algorithmic innovation over data preprocessing. The development of MTR algorithms has followed two main pathways: the adaptation of techniques from Multi-Label Classification (MLC) [14], and the design of novel architectures such as Multi-Objective Decision Trees [21] and Random Forests (RF) [15, 16]. The focus on model development has generated a gap in the literature: the dedicated study of feature selection tailored specifically to enhance MTR performance [17]. Authors in [18] confirmed this trend, highlighting that algorithm development remains the primary focus in the field. This is further demonstrated by a summary of seminal MTR papers, provided in Table I.

TABLE I. SUMMARY OF MTR MODEL DEVELOPMENT FOCUSED ON KEY LITERATURE

Paper	Core contribution	Limitations
MLC methods for MTR [14]	Adaptation of MLC algorithms for MTR tasks.	Does not address the "curse of dimensionality."
MTR via input space expansion [19]	A novel method to improve MTR performance by treating targets as inputs.	Model complexity and dimensionality issues are not discussed.
Feature selection for semi-supervised MTR [20]	A GA-based proposal for semi-supervised MTR.	Applied only to data with few labeled instances; not a general solution.
Prediction of drug efficacy using MTR [3]	Application of MTR models for ranking drug effectiveness.	Does not discuss feature selection as a mechanism for performance improvement.

As presented in Table I, even ground-breaking MTR papers focus on algorithm development, often overlooking the performance bottleneck created by high-dimensional feature spaces. To address this gap, this research proposes a systematic framework for feature selection. Methodologies for feature selection are generally divided into three categories: filter, wrapper, and embedded approaches [11, 26, 27]. The current work utilizes a wrapper method, which employs the underlying machine learning algorithm to evaluate candidate feature subsets, ensuring that the selected features are optimized for the specific predictive model [23].

The present study proposes leveraging GAs as a powerful, metaheuristic engine for this wrapper framework [24]. GAs, initially introduced in [25], are well-suited for this complex optimization task. Their evolutionary mechanisms of crossover and mutation enable an effective balance between exploring the vast search space of feature combinations and exploiting promising solutions [26], making them robust for identifying near-optimal feature subsets without succumbing to the local optima that can trap greedy algorithms [14, 15]. While the current study focuses on GAs for feature selection, other advanced techniques, such as Generative Adversarial Networks (GANs), have also demonstrated strong predictive performance in related tasks like ozone forecasting [27].

This research, therefore, develops and evaluates a GA-based feature selection methodology specifically designed for MTR tasks, with a focus on predicting multiple ozone metrics from the high-dimensional Environmental AQ-Bench Dataset. The core objective of the study is to identify optimal feature subsets that not only improve predictive accuracy but also enhance model interpretability and efficiency [22]. To achieve this goal, the present study makes several key contributions:

- Providing novel methodologies for MTR feature selection: The study introduces three GA-based methodologies. GA-FS-MTR-SO (Single-Objective) and GA-FS-MTR-MO (Multi-Objective) to identify a common feature subset for all targets, ideal for leveraging interdependencies, and GA-FS-TARGET to identify a unique feature subset for each target, offering flexibility when targets have differing drivers.
- Demonstrating performance improvement and dimensionality reduction: the proposed methods successfully address the "curse of dimensionality" by achieving a significant feature space reduction of up to 61.6%. This reduction in model complexity is accomplished while simultaneously enhancing predictive accuracy, with demonstrated performance improvements of up to 10.23% over the baseline Stacked Single Target (SST) method.
- Enhancing model interpretability: by systematically identifying the most influential features, the proposed approach transforms complex, high-dimensional models into more efficient and transparent ones [10], facilitating a clearer understanding of the model's decision-making process.
- Comprehensive experimental validation: The findings are supported by experiments on the publicly available AQ-Bench benchmark dataset. The effectiveness of the proposed methods is validated against baseline algorithms such as Single Target (ST), Stacked Single Target (SST), and Ensemble of Regressor Chains (ERC), and substantiated with statistical significance tests, such as the Friedman test [28] and p-values, providing robust evidence of their superiority.

II. METHODOLOGY

The proposed methodology integrated GAs with MTR models in a wrapper-based feature selection approach. The complete workflow was structured into four primary stages, as illustrated in Figure 1: (1) Data Preprocessing, (2) Feature Selection, (3) GA-based Model Training and Optimization, and (4) Model Testing.

A. Dataset and Preprocessing

The present study utilized the public AQ-Bench dataset [29], which contained global air quality measurements from over 5,500 stations (2010-2014) from the Tropospheric Ozone Assessment Report (TOAR) database. The initial preprocessing stage, shown as Step 1 in Figure 1, involved handling missing values and normalizing all features using a MinMaxScaler. The prepared dataset was then partitioned into a training set (70%), a validation set (10%), and a testing set (20%) for final, unbiased evaluation.

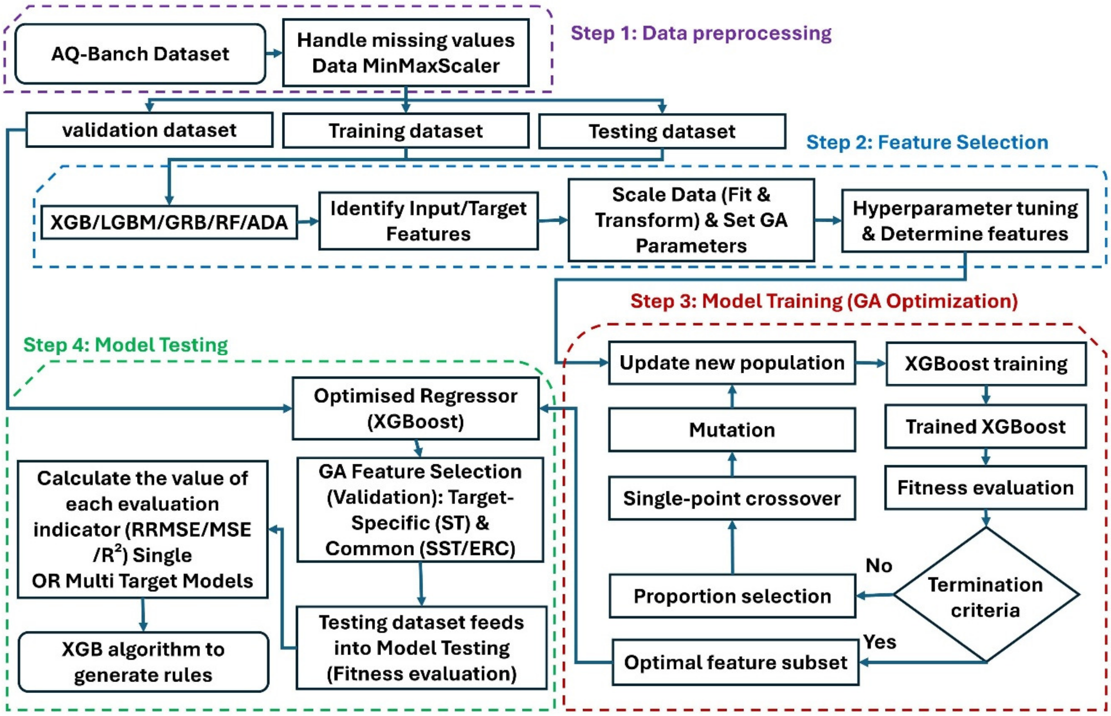

Fig. 1. Systematic workflow of the proposed GA-MTR methodology, illustrating the four main stages: data preprocessing, feature selection, GA-based model training, and model testing.

B. GA-Based Feature Selection Framework

A GA was employed to navigate the high-dimensional feature space. In this wrapper-based method, the performance of a regression model served as the guiding metric for identifying an optimal feature subset. The general process for this approach is displayed in Figure 2.

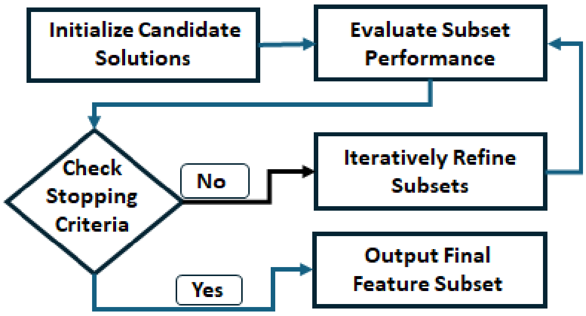

Fig. 2. Iterative process of a wrapper-based feature selection approach.

The core components of the GA were implemented as follows:

1) Encoding and Fitness Function

Each potential feature subset was encoded as a binary chromosome, where a '1' signified a feature's inclusion and a '0' its exclusion. The "fitness" of each chromosome was determined by training a regression model (e.g., XGBoost) on the corresponding feature subset and evaluating its predictive error. A lower error indicated higher fitness. The evaluation relied on two key metrics: Relative Root Mean Squared Error (RRMSE) for single-target evaluation and Average RRMSE (ARRMSE) to evaluate performance across all targets simultaneously.

$$\text{RRMSE} = \sqrt{\frac{\sum_{j=1}^{m}(y_j-\hat{y}_j)^2}{\sum_{j=1}^{m}(y_j-\bar{Y})^2}} \tag{1}$$

$$\text{ARRMSE} = \frac{1}{t}\sum_{k=1}^{t} RRMSE_k \tag{2}$$

2) Genetic Operators

The evolutionary process was driven by several operators, as seen in Step 3 in Figure 1. Tournament selection was used to choose parent chromosomes from the population based on their fitness. These parents then generated offspring through mechanisms such as single-point crossover, which combined genetic material, as illustrated in Figure 3. To maintain diversity, mutation was subsequently used to introduce small, random changes into these new child solutions. Finally, elitism was employed to ensure that the best-performing solution from each generation was preserved. This cycle was repeated until a termination criterion was met.

3) GA Hyperparameter Configuration

The GA was configured with a Population Size of 50 individuals for a maximum of 50 generations. A crossover probability of 0.8 and a mutation probability of 0.01 were used. Tournament selection size was set to 3, and elitism was set to preserve the 1-2 best solutions from each generation. Early stopping was implemented to halt the algorithm after 10-15 generations without any improvement in fitness. The fitness function was defined as the ARRMSE across all targets for the GA-FS-MTR-SO/MO models, and as the RRMSE for the GA-FS-TARGET model.

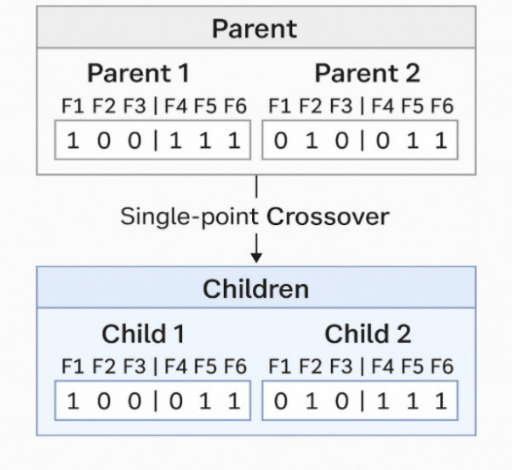

Fig. 3. Illustration of the single-point crossover operator, a primary mechanism for generating new candidate solutions from existing ones.

C. Proposed Feature Selection Strategies

This research introduced and evaluated three distinct GA-based feature selection strategies.

1) GA-FS-MTR-SO (Single-Objective)

This method sought a single, common feature subset by minimizing a single objective: the ARRMSE across all targets. Its implementation is detailed in Algorithm 1.

```
ALGORITHM: I. GA FS MTR SO
Input: X, Y, N, G_max, p_c, p_m
Output: S*
1. Initialize P ← {c_i ∈ {0,1}^p | i=1…N}
2. For g = 1…G_max:
3.   For each c in P:
4.     Compute ARRMSE(c) over all targets via ST models
5.     fitness(c) ← 1/(1 + ARRMSE(c))
6.   P_parents ← TournamentSelect(P, fitness)
7.   P_new ← Ø
8.   While |P_new| < N:
9.     p1,p2 ← RandomPair(P_parents)
10.    If rand()<p_c: (o1,o2)=Crossover(p1,p2) else o1=p1,o2=p2
11.    o1←Mutate(o1,p_m); o2←Mutate(o2,p_m)
12.    P_new ← P_new ∪ {o1,o2}
13.  P ← Elitism(P,P_new)
14.  Update best c*
15. Return S* ← indices of 1’s in best c*
```

2) GA-FS-MTR-MO (Multi-Objective)

This approach simultaneously minimized multiple objectives: the RRMSE for each target and the feature subset size. It used a non-dominated sorting algorithm to identify a Pareto front of optimal trade-off solutions, as described in Algorithm 2.

```
ALGORITHM: II. GA FS MTR MO
Input: X, Y, N, G_max, p_c, p_m
Output: Pareto front P*
1. Initialize P ← {c_i ∈ {0,1}^p}
2. For g = 1…G_max:
3.   For each c in P:
4.     For j=1…t: compute RRMSE_j(c)
5.     Objectives ← {RRMSE_1(c),…,RRMSE_t(c), |c|}
6.   P ← NonDominatedSortAndCrowding(P, Objectives)
7.   P_parents ← TournamentSelect(P, crowding-aware)
8.   Generate offspring as in Algorithm 1
9. Return P* ← final non-dominated set
```

3) GA-FS-TARGET (Per-Target)

In contrast, this strategy ran the GA independently for each target to find a unique, specialized feature subset. This approach was ideal for problems with heterogeneous targets, as outlined in Algorithm 3.

```
ALGORITHM: III. GA FS TARGET
Input: X, {y_j}, N, G_max, p_c, p_m
Output: {S_j*}
For each target j=1…t:
   1. Initialize P ← {c_i ∈ {0,1}^p}
   2. For g=1…G_max:
      3. For each c in P:
         4. Compute RRMSE_j(c)
         5. fitness(c) ← 1/(1 + RRMSE_j(c))
      6. P_parents ← TournamentSelect(P, fitness)
      7. Generate offspring as in Algorithm 1
      8. Update best c*_j
   9. S_j* ← indices of 1’s in c*_j
Return {S_1*,…,S_t*}
```

D. Experimental Setup and Evaluation

The final phase, as shown in Step 4, Figure 1, involved training and testing MTR models using the feature subsets identified by the GA strategies.

1) MTR Evaluation Models

The quality of the selected features was assessed using three established MTR algorithms: ST, which built independent models; SST, a two-layer meta-learning approach; and ERC, which modeled inter-target dependencies.

2) *Evaluation Protocol and Statistical Analysis*

The ST, SST, and ERC models were trained on the 70% training set using the selected features, with the final performance measured on the 20% held-out test set. Key metrics included ARRMSE, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). To ensure robust conclusions, statistical significance was assessed using paired t-tests (p-values), while the Friedman test with post-hoc analysis provided a comparative ranking of all methods.

III. RESULTS

The results present and interpret the empirical findings from the application of the proposed GA-based feature selection methodologies.

A. *Performance of Single-Objective Common Feature Subset (GA-FS-MTR-SO)*

To assess the efficacy of a single, optimized feature subset, the GA-FS-MTR-SO method was evaluated. The purpose of this experiment was to compare the ARRMSE of baseline MTR models against their performance when using the GA-selected features. The results of this comparison are presented in Tables II-IV. A simple statistical analysis was used to determine the significance of these performance differences. T-tests were used to analyze the relationship between the baseline and proposed methods, with the resulting p-values outlined in Table V.

TABLE II. ARRMSE COMPARISON: BASELINE ST VS. SINGLE-OBJECTIVE GA-FS-MTR(ST)-SO

Regressor	Baseline ST	GA-FS-MTR(ST)-SO
RF	0.41943	0.42661
XGB	0.44919	0.42471
ADA	0.44703	0.43719
GBM	0.42703	0.42950
LGBM	0.42696	0.42950

TABLE III. ARRMSE COMPARISON: BASELINE SST VS. SINGLE-OBJECTIVE GA-FS-MTR(SST)-SO

Regressor	Baseline SST	GA-FS-MTR(SST)-SO
RF	0.49394	0.48394
XGB	0.50721	0.45533
ADA	0.44952	0.43578
GBM	0.45624	0.45835
LGBM	0.45858	0.45835

To provide a holistic, statistical ranking of all methods, a Friedman test was conducted. The resulting Critical Distance (CD) diagrams for the Single-Objective and Multi-Objective approaches are shown in Figures 4 and 5, respectively. These diagrams should be interpreted as: lower ranks, positioned on the left, indicate superior performance. Methods connected by a horizontal bar are considered statistically similar in performance, while any gap between methods larger than the CD value is statistically significant.

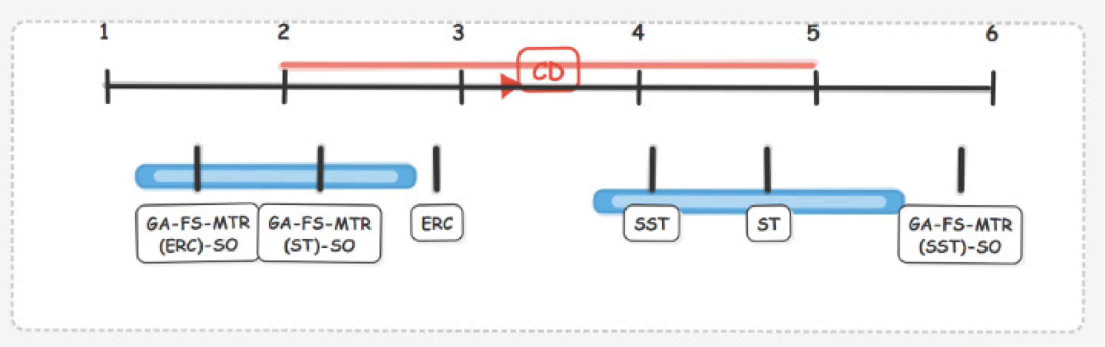

Fig. 4. CD diagram comparing Single-Objective (GA-FS-MTR-SO) methods against baseline methods (α =0.1).

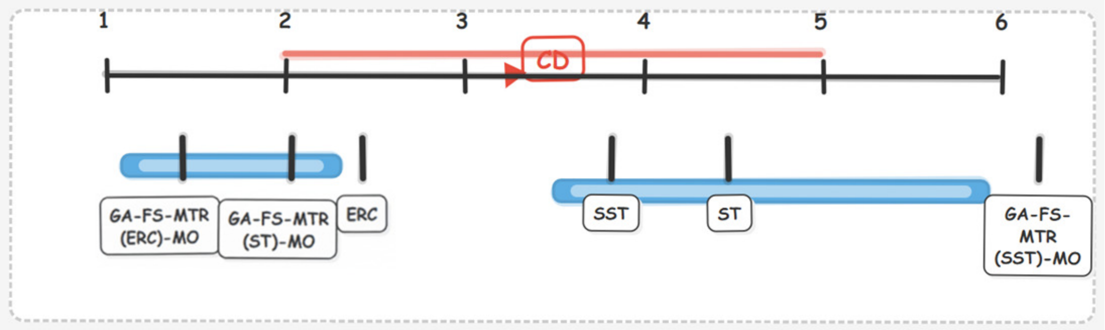

Fig. 5. CD diagram comparing Multi-Objective (GA-FS-MTR-MO) methods against baseline methods (α =0.1).

In Figures 4 and 5, a clear hierarchy was observed. The proposed GA-FS-MTR methods were found to consistently outperform baseline approaches (SST and ST), and the GA-FS-MTR(ERC) variants were found to win in both scenarios.

TABLE IV. ARRMSE COMPARISON: BASELINE ERC VERSUS SINGLE-OBJECTIVE GA-FS-MTR(ERC)-SO

Regressor	Baseline ERC	GA-FS-MTR(ERC)-SO
RF	0.50795	0.49795
XGB	0.86551	0.86873
ADA	0.49230	0.51582
GBM	0.50144	0.62750
LGBM	0.55082	0.62750

TABLE V. STATISTICAL SIGNIFICANCE (P-Values) FOR SINGLE-OBJECTIVE FEATURE SELECTION VERSUS BASELINE MODELS

Regressor	p-value (baseline)			p-value (GA-FS-MTR (ST)-SO		
	ST	SST	ERC	ST	SST	ERC
RF	0.0088	0.0309	0.0306	0.0000	0.0236	0.0000
XGB	0.0000	0.0000	0.00185	0.0357	0.0446	0.0719
ADA	0.0087	0.0005	0.02066	0.0489	0.0089	0.0563
GBM	0.0000	0.0614	0.02419	0.0000	0.0000	0.0002
LGBM	0.0264	0.0617	0.07603	0.0147	0.0572	0.0305

TABLE VI. ARRMSE COMPARISON: BASELINE ST VERSUS MULTI-OBJECTIVE GA-FS-MTR(ST)-MO

Regressor	Baseline ST	GA-FS-MTR(ST)-MO
RF	0.41943	0.42016
XGB	0.43293	0.43044
ADA	0.44305	0.44549
GBM	0.42827	0.42608
LGBM	0.42776	0.42653

B. *Performance of Multi-Objective Common Feature Subset (GA-FS-MTR-MO)*

The next set of analysis aimed to evaluate the Multi-Objective (MO) approach. The ARRMSE comparisons are presented in Tables VI-VIII, with corresponding significance tests depicted in Table IX.

C. *Performance of the Per-Target Approach (GA-FS-TARGET)*

To assess the viability of a target-specific strategy, the GA-FS-TARGET method was employed. The results are shown in Tables X-XII.

TABLE VII. ARRMSE COMPARISON: BASELINE SST VERSUS MULTI-OBJECTIVE GA-FS-MTR(SST)-MO

Regressor	Baseline SST	GA-FS-MTR(SST)-MO
RF	0.45974	0.46252
XGB	0.48166	0.47832
ADA	0.45066	0.45264
GBM	0.46380	0.43395
LGBM	0.50014	0.46187

TABLE VIII. ARRMSE COMPARISON: BASELINE ERC VERSUS MULTI-OBJECTIVE GA-FS-MTR(ERC)-MO

Regressor	Baseline ERC	GA-FS-MTR(ERC)-MO
RF	0.48570	0.47930
XGB	0.87334	0.91671
ADA	0.52673	0.51363
GBM	0.45847	0.49675
LGBM	0.56043	0.53655

D. *Impact on Feature Dimensionality and Overall Performance*

Changes in feature set size and overall performance were compared. Table XIII quantifies the feature reduction, while Tables XIV-XVI summarize the percentage change in ARRMSE.

TABLE IX. STATISTICAL SIGNIFICANCE (P-VALUES) FOR MULTI-OBJECTIVE FEATURE SELECTION VERSUS BASELINE MODELS

Regressor	p-value (baseline)			p-value (GA-FS-MTR(ST)-MO		
	ST	SST	ERC	ST	SST	ERC
RF	0.0184	0.0246	0.0000	0.00001	0.0236	0.0000
XGB	0.0000	0.0594	0.0015	0.03571	0.0446	0.0719
ADA	0.0133	0.0194	0.0036	0.04896	0.0089	0.0563
GBM	0.0000	0.0000	0.0064	0.0000	0.0000	0.0002
LGBM	0.0603	0.0196	0.2177	0.0187	0.0597	0.1704

TABLE X. ARRMSE COMPARISON: BASELINE MTR MODELS VERSUS PER-TARGET METHOD (GA-FS-TARGET)

Regressor	Baseline-TARGET-ST	GA-FS-TARGET-ST	Baseline-TARGET-SST	GA-FS-TARGET-SST	Baseline-TARGET-ERC	GA-FS-TARGET-ERC
RF	0.41943	0.42068	0.44299	0.45299	0.42656	0.43656
XGB	0.45979	0.42958	0.49818	0.46818	0.47270	0.44270
ADA	0.43426	0.43676	0.45147	0.45347	0.43189	0.43389
GBM	0.45787	0.44335	0.44045	0.48045	0.44447	0.43447
LGBM	0.43120	0.42629	0.46895	0.45895	0.45090	0.44090

TABLE XI. RRMSE PERFORMANCE FOR SPECIFIC TARGET 'O3_DAYTIME_AVG' USING GA-FS-TARGET METHOD

Regressor	RRMSE for O3_daytime_avg
RF	0.236742
XGB	0.173338
ADA	0.167670
GBM	0.172462
LGBM	0.187553

TABLE XII. STATISTICAL SIGNIFICANCE (P-VALUES) FOR THE PER-TARGET METHOD (GA-FS-TARGET) VERSUS BASELINE MODELS

Regressor	ST	SST	ERC
RF	0.05248	0.01041	0.06221
XGB	0.06575	0.03514	0.05161
ADA	0.00000	0.06402	0.00000
GBM	0.06235	0.06475	0.06080
LGBM	0.00001	0.06108	0.03007

TABLE XIII. FEATURE SET REDUCTION ACHIEVED BY GA-FS-MTR (SO AND MO) APPROACHES

Regressor	Total Features	GA-FS-MTR-SO	Reduction (%)	GA-FS-MTR-MO	Reduction (%)
RF	100	47	47%	56	44%
XGB	100	58	42%	50	50%
ADA	100	50	50%	49	51%
GBM	125	69	44.8%	48	61.6%
LGBM	125	65	48%	51	59.2%

TABLE XIV. PERCENTAGE CHANGE IN ARRMSE VERSUS BASELINE ST

Regressor	GA-FS-MTR-SO (ST)	GA-FS-MTR-MO (ST)	GA-FS-TARGET-ST
RF	-1.71%	-0.17%	-0.30%
XGB	+5.45%	+0.58%	+6.57%
ADA	+2.20%	-0.55%	-0.58%
GBM	-0.58%	+0.51%	+3.17%
LGBM	-0.59%	+0.29%	+1.14%

TABLE XV. PERCENTAGE CHANGE IN ARRMSE VERSUS BASELINE SST

Regressor	GA-FS-MTR-SO (SST)	GA-FS-MTR-MO (SST)	GA-FS-TARGET-SST
RF	+2.02%	-0.60%	-2.26%
XGB	+10.23%	+0.69%	+6.02%
ADA	+3.06%	-0.44%	-0.44%
GBM	-0.46%	+6.44%	-9.08%
LGBM	+0.05%	+7.65%	+2.13%

TABLE XVI. PERCENTAGE CHANGE IN ARRMSE VERSUS BASELINE ERC

Regressor	GA-FS-MTR-SO (ERC)	GA-FS-MTR-MO (ERC)	GA-FS-TARGET-ERC
RF	+1.97%	+1.32%	-2.34%
XGB	-0.37%	-4.97%	+6.35%
ADA	-4.78%	+2.49%	-0.46%
GBM	-25.14%	-8.35%	+2.25%
LGBM	-13.92%	+4.26%	+2.22%

IV. DISCUSSION

The results of the present study show that GA-based feature selection can significantly enhance MTR performance. The most obvious finding to emerge from the analysis is the dual benefit of this approach, since it provides reduced model complexity and, in many cases, enhanced predictive accuracy. As presented in Table XIII, the substantial feature reductions, reaching up to 61.6%, directly address the "curse of dimensionality," a well-documented challenge in machine learning. This indicates that the GA effectively acts as a filter, removing noisy and redundant features. To contextualize these findings and validate the efficacy of the proposed approach, the latter's performance was compared against results from several studies, as presented in Table XVII.

This comparative analysis reveals that the proposed method achieves state-of-the-art performance, delivering substantial feature reductions that are competitive with those reported in the literature. The proposed method demonstrates exceptional scalability by succeeding on the high-dimensional AQ-Bench dataset (100+ features), which presents a far more complex challenge than the smaller-scale datasets used in many benchmark studies. Furthermore, the results of the present study strongly corroborate the findings in [30, 31], confirming that MO optimization is a superior strategy for achieving more efficient dimensionality reduction in MTR contexts. This validation establishes the proposed framework as a potent and effective methodology for tackling complex, real-world environmental modeling tasks.

Another important finding is that no universally superior feature selection strategy exists. The GA-FS-MTR approaches are well-suited for problems where targets are strongly correlated, while the GA-FS-TARGET approach demonstrates its unique value in situations requiring precision for heterogeneous targets. This provides a practical contribution, as it offers a framework for practitioners to choose a strategy based on the problem's domain context.

Perhaps the most important result observed, however, was a critical interaction between the feature selection process and the MTR model's architecture. As presented in Table XVI, the marked performance degradation of the ERC model in certain cases, despite its top ranking in the Friedman tests, as evidenced in Figures 4 and 5, is highly instructive. The findings suggest that the ERC model's chained-prediction architecture creates a strong, sequential interdependency. It can be hypothesized that the GA, by optimizing for overall ARRMSE, removed features that were not top predictors for a single target but were nevertheless vital as "informational bridges" in the predictive chain. The removal of these features likely made the model's performance brittle. An alternative interpretation could be that the GA's fitness landscape for ERC is more prone to local optima.

This observation holds significant implications and encourages critical thought about future research directions. It suggests that standard fitness functions may be insufficient for structurally complex models like ERC. Future research should explore more sophisticated fitness criteria that reward the preservation of structural linkage. This finding serves as a cautionary note: practitioners cannot treat feature selection as a black-box preprocessing step, but must instead consider it in combination with the specific mechanics of the chosen algorithm.

In conclusion, this research validates GA-based feature selection as a potent methodology for improving MTR. It moves beyond a simple demonstration of efficacy to provide a strategic framework for practitioners and uncovers the critical need to co-design feature selection and modeling strategies for structurally intricate algorithms.

TABLE XVII. COMPARATIVE FEATURE REDUCTION PERFORMANCE OF GA-MTR METHODOLOGY ACROSS KEY STUDIES

Study	Identifier	Total features	Selected features (SO)	Reduction % (SO)	Selected features (MO)	Reduction % (MO)
Proposed method	RF	100	47	47%	56	44%
	XGB	100	58	42%	50	50%
	ADA	100	50	50%	49	51%
	GBM	125	69	44.8%	48	61.6%
	LGBM	125	65	48%	51	59.2%
[30]	CART-Color	12	8	33%	6	50%
	RF-Color	12	7	42%	5	58%
	SVM-Color	12	9	25%	7	42%
	Benchmarks (Avg)	15.2	8.4	45%	7.1	53%
[31]	Air quality dataset	42	18	57%	15	64%
	Pollution monitoring	38	16	58%	13	66%
[32]	andro	30	10	67%	10	67%
	edm	16	7	56%	9	44%
	enb	8	4	50%	5	38%
	jura	15	8	47%	8	47%
	scpf	23	12	48%	8	65%
	sf1	10	3	70%	2	80%
	sf2	10	3	70%	1	90%
	slump	7	3	57%	3	57%
[33]	ERC-Dataset1	50	23	54%	18	64%
	ERC-Dataset2	45	21	53%	16	64%
[19]	SST-Forest	16	9	44%	8	50%
	ERC-Water	16	8	50%	6	63%
	SST-Energy	8	4	50%	3	63%
[34]	Cross-Stack-O3	25	12	52%	9	64%
	Ensemble-VOCs	35	18	49%	14	60%
[35]	Ensemble-MTR-18DS	22	11	50%	8	64%

V. CONCLUSION

The present study focused on enhancing the prediction of multiple ozone metrics from the high-dimensional AQ-Bench Dataset by developing and evaluating a suite of Genetic Algorithm (GA)-based feature selection methodologies for Multi-Target Regression (MTR). The investigation confirmed that the proposed GA-based approaches provide a dual benefit: they significantly reduce model complexity while simultaneously improving predictive performance. The research has shown that feature dimensionality was successfully reduced by up to 61.6%, and the Friedman tests demonstrated that the proposed GA-FS-MTR methods consistently and significantly outperformed standard baseline approaches.

The primary contribution of this work is the validation of a strategic framework for applying feature selection to MTR problems. There is no single superior strategy; instead, the optimal approach is contingent on the nature of the prediction task. The GA-FS-MTR methods (both single- and multi-objective) are best suited for problems with correlated targets, whereas the GA-FS-TARGET approach provides essential precision when targets are heterogeneous. A key finding, however, was the observed performance degradation of the structurally complex Ensemble of Regressor Chains (ERC) model under certain conditions, suggesting that its chained architecture is highly sensitive to the removal of features that may serve as informational bridges. This observation constitutes a principal limitation and a critical insight of this study.

Future research should be directed toward developing more sophisticated fitness functions that are aware of a model's underlying architecture, potentially rewarding the preservation of structural linkage features. Furthermore, investigating hybrid methods that combine the proposed wrapper-based approach with computationally efficient filter techniques could yield further improvements. In conclusion, this research validates GA-based feature selection as an effective methodology for optimizing MTR models and provides practitioners with a clear, evidence-based guide for its application in complex environmental modeling tasks.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York City, NY, USA: Springer New York, 2009.

[2] D. Kocev, S. Džeroski, M. D. White, G. R. Newell, and P. Griffioen, "Using Single- and Multi-target Regression Trees and Ensembles to Model a Compound Index of Vegetation Condition," *Ecological Modelling*, vol. 220, no. 8, pp. 1159–1168, Apr. 2009, https://doi.org/10.1016/j.ecolmodel.2009.01.037.

[3] G. R. Brindha, B. S. Rishiikeshwer, B. Santhi, K. Nakendraprasath, R. Manikandan, and A. H. Gandomi, "Precise Prediction of Multiple Anticancer Drug Efficacy using Multi Target Regression and Support Vector Regression Analysis," *Computer Methods and Programs in Biomedicine*, vol. 224, Sept. 2022, Art. no. 107027, https://doi.org/10.1016/j.cmpb.2022.107027.

[4] S. Barbon Junior *et al.*, "Multi-Target Prediction of Wheat Flour Quality Parameters With Near Infrared Spectroscopy," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 342–354, Jun. 2020, https://doi.org/10.1016/j.inpa.2019.07.001.

[5] B. F. Darst, K. C. Malecki, and C. D. Engelman, "Using Recursive Feature Elimination in Random Forest to Account for Correlated Variables in High Dimensional Data," *BMC Genetics*, vol. 19, no. S1, Sept. 2018, Art. no. 65, https://doi.org/10.1186/s12863-018-0633-8.

[6] Z. Wen and Y. Li, "A Spatial-Constrained Multi-target Regression Model for Human Brain Activity Prediction," *Applied Informatics*, vol. 3, no. 1, Dec. 2016, Art. no. 10, https://doi.org/10.1186/s40535-016-0026-x.

[7] F. I. Lewis and M. P. Ward, "Improving Epidemiologic Data Analyses Through Multivariate Regression Modelling," *Emerging Themes in Epidemiology*, vol. 10, no. 1, May 2013, Art. no. 4, https://doi.org/10.1186/1742-7622-10-4.

[8] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga, "A Survey on Multi-output Regression," *WIREs Data Mining and Knowledge Discovery*, vol. 5, no. 5, pp. 216–233, Sept. 2015, https://doi.org/10.1002/widm.1157.

[9] S. S. Du, J. D. Lee, H. Li, L. Wang, and X. Zhai, "Gradient Descent Finds Global Minima of Deep Neural Networks," in *36th International Conference on Machine Learning*, Long Beach, CA, USA, 2019, Art. no. 97, https://doi.org/10.48550/ARXIV.1811.03804.

[10] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, May 2019, https://doi.org/10.1038/s42256-019-0048-x.

[11] T. Aho, B. Ženko, S. Džeroski, and T. Elomaa, "Multi-Target Regression with Rule Ensembles," *Journal of Machine Learning Research*, vol. 13, pp. 2367–2407, Aug. 2012.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 1135–1144, https://doi.org/10.1145/2939672.2939778.

[13] P. Domingos, "A Few Useful Things to Know About Machine Learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012, https://doi.org/10.1145/2347736.2347755.

[14] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-Labelsets for Multilabel Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079–1089, Jul. 2011, https://doi.org/10.1109/TKDE.2010.164.

[15] D. Di Fina, S. Karaman, A. D. Bagdanov, and A. Del Bimbo, "MORF: Multi-Objective Random Forests for face characteristic estimation," in *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Karlsruhe, Germany, Aug. 2015, pp. 1–6, https://doi.org/10.1109/AVSS.2015.7301793.

[16] D. Kocev, C. Vens, J. Struyf, and S. Džeroski, "Ensembles of Multi-Objective Decision Trees," in *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R. L. D. Mantaras, S. Matwin, D. Mladenič, and A. Skowron, Eds. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2007, vol. 4701, pp. 624–631.

[17] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003.

[18] M. Petković, D. Kocev, and S. Džeroski, "Feature Ranking for Multi-Target Regression," *Machine Learning*, vol. 109, no. 6, pp. 1179–1204, Jun. 2020, https://doi.org/10.1007/s10994-019-05829-8.

[19] E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, and I. Vlahavas, "Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs," *Machine Learning*, vol. 104, no. 1, pp. 55–98, Jul. 2016, https://doi.org/10.1007/s10994-016-5546-z.

[20] F. H. Syed, M. A. Tahir, M. Rafi, and M. D. Shahab, "Feature Selection for Semi-Supervised Multi-target Regression Using Genetic Algorithm," *Applied Intelligence*, vol. 51, no. 12, pp. 8961–8984, Dec. 2021, https://doi.org/10.1007/s10489-021-02291-9.

[21] A. L. Blum and P. Langley, "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 245–271, Dec. 1997, https://doi.org/10.1016/S0004-3702(97)00063-5.

[22] J. Li *et al.*, "Feature Selection: A Data Perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, Nov. 2018, https://doi.org/10.1145/3136625.

[23] R. Kohavi and G. H. John, "Wrappers for Feature Subset Selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, https://doi.org/10.1016/S0004-3702(97)00043-X.

[24] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019, https://doi.org/10.1109/ACCESS.2019.2906757.

[25] J. H. Holland, *Adaptation in Natural and Artificial Systems*, 1st ed. Cambridge, MA, USA: MIT Press, 1992.

[26] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug. 2016, https://doi.org/10.1109/TEVC.2015.2504420.

[27] S. Khedekar and S. Thakare, "Predicting Air Pollution Levels in Pune, India using Generative Adversarial Networks," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17405–17413, Oct. 2024, https://doi.org/10.48084/etasr.8512.

[28] M. Friedman, "A Comparison of Alternative Tests of Significance for the Problem of m Rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, Mar. 1940, https://doi.org/10.1214/aoms/1177731944.

[29] C. Betancourt, T. Stomberg, R. Roscher, M. G. Schultz, and S. Stadtler, "AQ-Bench: a Benchmark Dataset for Machine Learning on Global Air Quality Metrics," *Earth System Science Data*, vol. 13, no. 6, pp. 3013–3033, Jun. 2021, https://doi.org/10.5194/essd-13-3013-2021.

[30] A. S. Brar and K. Singh, "A Multi-Objective Stacked Regression Method for Distance Based Colour Measuring Device," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 5530, https://doi.org/10.1038/s41598-024-54785-4.

[31] S. Masmoudi, H. Elghazel, D. Taieb, O. Yazar, and A. Kallel, "A Machine-Learning Framework for Predicting Multiple Air Pollutants' Concentrations via Multi-target Regression and Feature Selection,"

Science of the Total Environment, vol. 715, May 2020, Art. no. 136991, https://doi.org/10.1016/j.scitotenv.2020.136991.

[32] F. H. Syed, M. A. Tahir, J. Frnda, M. Rafi, M. S. Anwar, and J. Nedoma, "Toward an Optimal and Structured Feature Subset Selection for Multi-Target Regression Using Genetic Algorithm," *IEEE Access*, vol. 11, pp. 121966–121977, 2023, https://doi.org/10.1109/ACCESS.2023.3327870.

[33] G. Melki, A. Cano, V. Kecman, and S. Ventura, "Multi-Target Support Vector Regression via Correlation Regressor Chains," *Information Sciences*, vol. 415–416, pp. 53–69, Nov. 2017, https://doi.org/10.1016/j.ins.2017.06.017.

[34] Z. Ning *et al.*, "Prediction and Explanation for Ozone Variability Using Cross-stacked Ensemble Learning Model," *Science of The Total Environment*, vol. 935, Jul. 2024, Art. no. 173382, https://doi.org/10.1016/j.scitotenv.2024.173382.

[35] O. Reyes, H. M. Fardoun, and S. Ventura, "An Ensemble-Based Method for the Selection of Instances in the Multi-target Regression Problem," *Integrated Computer-Aided Engineering*, vol. 25, no. 4, pp. 305–320, Sept. 2018, https://doi.org/10.3233/ICA-180581.