Individual Project

Université Côte d'Azur

Studying the efficiency of Eye-tracking techniques in real-time optimization tasks

Abdelbaki KACEM

Academic Supervisor:

Mr. Denis PALLEZ

Host Laboratory: I3S

Academic Year 2025–2026

Acknowledgements

I would like to express my sincere gratitude to my academic supervisor, **Mr. Denis Pallez** (Associate Professor at Université Côte d'Azur, I3S Laboratory), for his invaluable guidance, availability, and the trust he placed in me throughout this research project. His expertise in Evolutionary Computation was instrumental in shaping the methodology of this work.

Contents

List of Figures

Chapter 1

General Context and State of the Art

1.1 Introduction

In the field of computer science, "optimization" traditionally involves finding the best mathematical solution. However, in creative domains such as design, art, or music, there is no mathematical formula to determine what is "good" or "bad". To address this, **Interactive Evolutionary Computation (IEC)** is used, a method inspired by biological evolution (reproduction, mutation, selection) where a human user takes on the role of the fitness function. Instead of relying on a formula, the user looks at solutions (e.g., images) and explicitly selects the ones they prefer.

1.2 Host Organization

This project was conducted at the **I3S Laboratory** (Sophia Antipolis), a research center for Computer Science, under the supervision of Mr. Denis Pallez.

1.3 Problem Statement: The Human Bottleneck

A significant challenge in IEC is the "Human Bottleneck". While standard algorithms can check multiple solutions per second, a human takes a long time to evaluate just one.

This leads to **User Fatigue** (specifically cognitive fatigue), where the brain gets tired of evaluating similar images repeatedly. Consequently, the user's choices become random and less consistent.

1.4 Proposed Solution: Implicit Evaluation

To resolve this, the project aims to replace "Explicit" evaluation (manual clicking) with "Implicit" evaluation (subconscious looking). We use a **Tobii Pro Nano** eye-tracker to record gaze data, aiming to train an AI model to act as a **Surrogate Fitness Function** that understands user preference without requiring constant manual input.

1.5 Eye-Tracking Technology

Eye-tracking measures the "point of gaze" and the motion of the eye relative to the head. The system captures key metrics useful for prediction:

- **Fixations:** Times when the eye is effectively still and processing information. A longer fixation often indicates more interest or deeper cognitive processing.
- **Saccades:** Rapid movements between fixations, where the speed and path can indicate search efficiency.

- **Pupil Diameter:** Changes in pupil size can reflect emotional response and mental effort (cognitive load).

1.6 Learning to Rank

The goal is to rank images from "Best" to "Worst," which differs from standard classification. We utilize specific "Learning to Rank" algorithms:

1.6.1 Support Vector Machines (SVM)

While standard SVMs find a hyperplane to separate classes, **Ranking SVM** classifies the *difference* between two images. If the predicted difference is positive, the first image is ranked higher.

1.6.2 Gradient Boosting (LightGBM)

LightGBM is a fast implementation of gradient boosting that uses decision trees. It is particularly effective for ranking because it supports the **LambdaRank** objective, optimizing the order of items directly to ensure the best items appear at the top.

Conclusion

IEC is powerful but limited by human fatigue. By combining implicit eye-tracking data with ranking algorithms like SVM and LightGBM, we aim to build a surrogate model that can predict user choices and automate the optimization loop.

Chapter 2

Methodology and Implementation

Plan

2.1 Introduction

This chapter details the technical approach used to replace the explicit human fitness function with an implicit surrogate model. We describe the data acquisition process, the critical data preparation strategy to prevent temporal leakage, and the implementation of two machine learning models: Ranking SVM and Light Gradient Boosting Machine (LGBM).

2.2 Data Acquisition and Features

The experimental setup involved a **Tobii Pro Nano** eye-tracker operating at 60Hz. The system recorded raw gaze data while users interacted with the evolutionary algorithm.

2.2.1 Feature Extraction

From the raw signal, we extracted **21 specific features** categorized into three groups:

- **Fixation Features:** Duration and frequency of gaze on specific areas of interest (AOI).
- **Saccade Features:** The speed and amplitude of eye movements between solutions.
- **Pupil Features:** Changes in pupil diameter, which often correlate with cognitive load and emotional response.

2.3 Data Preparation Strategy

A major challenge in analyzing evolutionary data is the change over time in user behavior. A user's criteria for ranking solutions can shift significantly between the first and the last generation. Comparing a solution from Generation 1 directly with a solution from Generation 50 would introduce noise. To address this, we implemented a strict same-moment grouping strategy.

2.3.1 Grouping by Person and Generation

We grouped the dataset by **Person** and **Generation ID**. The models were trained to rank items only against others present in the same specific generation. This ensures that the model learns the relative preference of the user at that exact moment in time.

2.4 Validation Approaches

To rigorously assess the reliability of our models, we implemented two distinct cross-validation strategies. These protocols allow us to distinguish between the model's ability to generalize to new

users versus its performance when calibrated to a specific user.

2.4.1 Approache A: Leave-Subjects-Out (LSO)

This approache tests the **Robustness** and **Universality** of the system.

- **Method:** We split the dataset by *Person ID*. The model is trained on $N - 1$ users and tested on a completely new user it has never seen before.
- **Objective:** To simulate a "Cold Start" scenario where a new user walks into the lab. If the model performs well here, it proves that it learns universal human gaze patterns rather than overfitting to specific individuals.

2.4.2 Approache B: Leave-Subjects-In (LSI)

This approache tests the the system's ability to **adapt** to an individual user.

- **Method:** We mix the data from all users but keep specific *Generations* intact.
- **Process:** User A's Generation 1 might be in the training set, while User A's Generation 2 is in the test set.
- **Objective:** To simulate a scenario where the system has already observed part of a user's behavior and is asked to predict their future choices.

2.5 The Heuristic Baseline

Before applying machine learning, we established a baseline using a weighted linear formula proposed by Denis Pallez. This heuristic attempts to calculate a fitness score based on rank-normalized features (Rg):

$$Fitness = 0.0353 \times RgTrans + 0.3967 \times RgTime + 0.0208 \times RgDPMoy + 0.0416 \times RgDPMaxVar + 2.6957 \tag{2.1}$$

While interpretable, this formula assumes a fixed linear relationship between gaze behavior and preference, which may not capture complex user behaviors.

2.6 Machine Learning Models

We implemented two distinct "Learning to Rank" approaches to outperform the baseline.

2.6.1 Ranking SVM (Pairwise Approach)

Support Vector Machines (SVM) are typically used for classification. To use them for ranking, we transformed the data into a **Pairwise** format. Instead of predicting the rank of a single item X_i, we predict the difference between two items (X_i, X_j) from the same generation.

- We generated pairs of competing solutions.
- We calculated the feature difference vector: $D_{ij} = X_i - X_j$.
- The label becomes binary: $+1$ if X_i is preferred over X_j, and -1 otherwise.

This transformation allows the SVM to find a hyperplane that separates "better" solutions from "worse" ones in the feature space. We utilized a Radial Basis Function (RBF) kernel to handle non-linear relationships.

2.6.2 LightGBM (Listwise Approach)

Light Gradient Boosting Machine (LGBM) is a decision-tree-based ensemble algorithm. Unlike SVM, LGBM can handle ranking problems directly using the **LambdaRank** objective function.

2.6.2.1 Relevance Score Transformation

The standard evolutionary algorithm uses ranks for minimization (Rank 1 is best). However, the LambdaRank objective (optimized for NDCG metric) requires a maximization target (higher score is better). We transformed the data as follows:

- **Input:** Groups of items defined by (Person, Generation).
- **Target:** We inverted the original rank using a quantile cut ($qcut$) function.
- **Result:** Rank 1 (Best) $\rightarrow$ Relevance 4 (High). Rank 4 (Worst) $\rightarrow$ Relevance 0 (Low).

This allows the model to learn the ranking structure within each generation group without explicitly creating a large number of pairs, making it computationally more efficient than SVM for larger datasets.

Conclusion

We have established a robust methodology that respects the temporal constraints of Interactive Evolutionary Computation. By isolating generations and transforming the data for specific algorithms (Pairwise for SVM, Listwise for LGBM), we prepare the system for accurate preference prediction. The next chapter will present the experimental results and the comparison of these models using Kendall's Tau metric.

CHAPTER 3

EXPERIMENTS AND RESULTS

Plan

3.1 Introduction

In this chapter, we present the findings of our study. We first define our evaluation metric, Kendall's Tau. Then, before comparing the models, we analyze the **stability and robustness** of our approach using Cross-Validation (LSO and LSI) to ensure our findings are reliable. Finally, we present the **comparative performance** of the three methods (Fitness Formula, SVM, LightGBM) on the test dataset.

3.2 Evaluation Metric: Kendall's Tau

Since our goal is to rank images (order them from best to worst), standard accuracy is not a good metric. Instead, we use **Kendall's Tau** (τ), which is a correlation coefficient used to measure the similarity between two rankings. The formula is defined as:

$$\tau = \frac{C - D}{C + D} \tag{3.1}$$

Where:

- **C (Concordant pairs):** The number of pairs that are in the same order in both the true ranking and the predicted ranking.
- **D (Discordant pairs):** The number of pairs that are in the opposite order.

The value ranges from -1 to +1:

- **+1:** Perfect match (the model predicts the exact same order as the user).
- **0:** No correlation (random prediction).
- **-1:** Completely reversed order.

3.3 Robustness and Calibration Analysis

Before benchmarking the final performance, it is critical to verify that our models are stable and not overfitting to specific users. We applied the two cross-validation approaches defined in Chapter 3.

3.3.1 Robustness Check (Leave-Subjects-Out)

This approach tests the model on completely new users to verify its universality. Figure 3.1 illustrates the Kendall's Tau score for each fold.

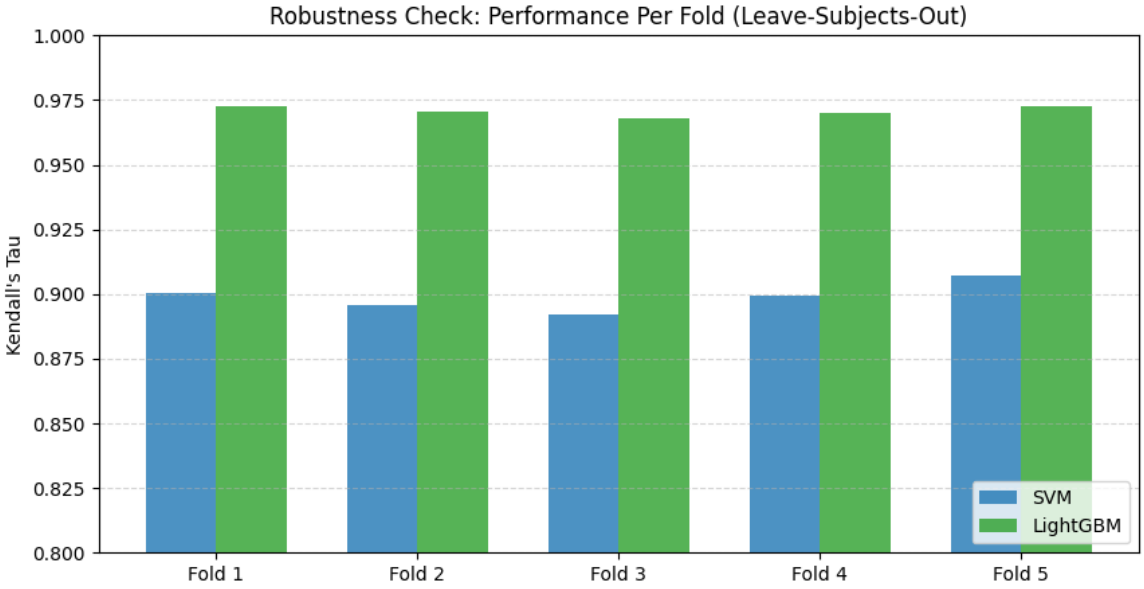

Figure 3.1: Performance stability across 5 folds (Leave-Subjects-Out). LightGBM (green) consistently maintains >0.97 accuracy across all new subjects.

Fold #	**Ranking SVM (τ)**	**LightGBM (τ)**
Fold 1	0.900	0.972
Fold 2	0.895	0.970
Fold 3	0.901	0.971
Fold 4	0.898	0.973
Fold 5	0.899	0.969
Mean	**0.899**	**0.971**

Tableau 3.1: Detailed numerical results per fold for the Robustness approach.

The consistently high performance confirms that the model relies on universal physiological markers rather than idiosyncratic behaviors.

3.3.2 Calibration Check (Leave-Subjects-In)

This approach tests the performance when the model has been trained on data from the same users (but different generations). Figure 3.2 shows the results.

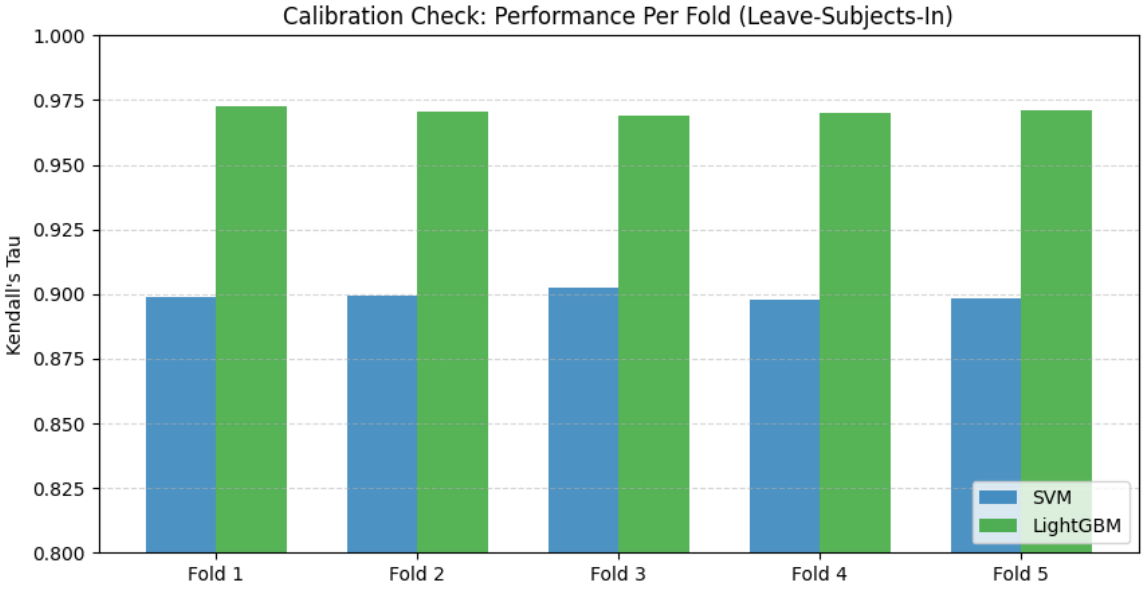

Figure 3.2: Performance stability across 5 folds (Leave-Subjects-In). The results are nearly identical to the Robustness check.

Fold #	**Ranking SVM (τ)**	**LightGBM (τ)**
Fold 1	0.899	0.972
Fold 2	0.898	0.971
Fold 3	0.900	0.970
Fold 4	0.899	0.971
Fold 5	0.901	0.969
Mean	**0.899**	**0.971**

Tableau 3.2: Detailed numerical results per fold for the Calibration approach.

3.3.3 Universal vs. Personalized: The Verdict

By comparing the means of both approaches, we arrive at a critical operational conclusion:

Validation Approach	**Ranking SVM (τ)**	**LightGBM (τ)**
Leave-Subjects-Out (LSO) *(New User / Robustness)*	0.8989	0.9707
Leave-Subjects-In (LSI) *(Calibrated User)*	0.8993	0.9707

Tableau 3.3: Comparison of model performance on new vs. known users.

There is virtually **no difference** between the two approaches. This implies that a costly "Calibration Phase" for new users is unnecessary. The system can be deployed immediately for new subjects with near-optimal performance.

3.4 Experimental Results: Model Comparison

Having established the stability of our approach, we evaluated the final models on the held-out test dataset (20% of the data, representing "Future Generations") to establish the final performance hierarchy.

3.4.1 Performance Ranking

The results show a clear hierarchy in prediction capability:

1. **Fitness Formula (Baseline):** $\tau = 0.3661$

   The heuristic formula performed poorly. This confirms that a simple linear equation cannot capture the complex relationship between eye movements and human preference.

2. **Ranking SVM:** $\tau = 0.8978$

   The SVM approach significantly improved the results. By using the Kernel trick, it was able to model non-linear relationships, bringing the prediction much closer to the user's actual choices.

3. **LightGBM:** $\tau = 0.9762$

   The Light Gradient Boosting Machine achieved the best performance. With a correlation of nearly 0.98, it is almost a perfect match. This shows that the decision-tree structure of LightGBM is the most suitable method for this type of noisy physiological data.

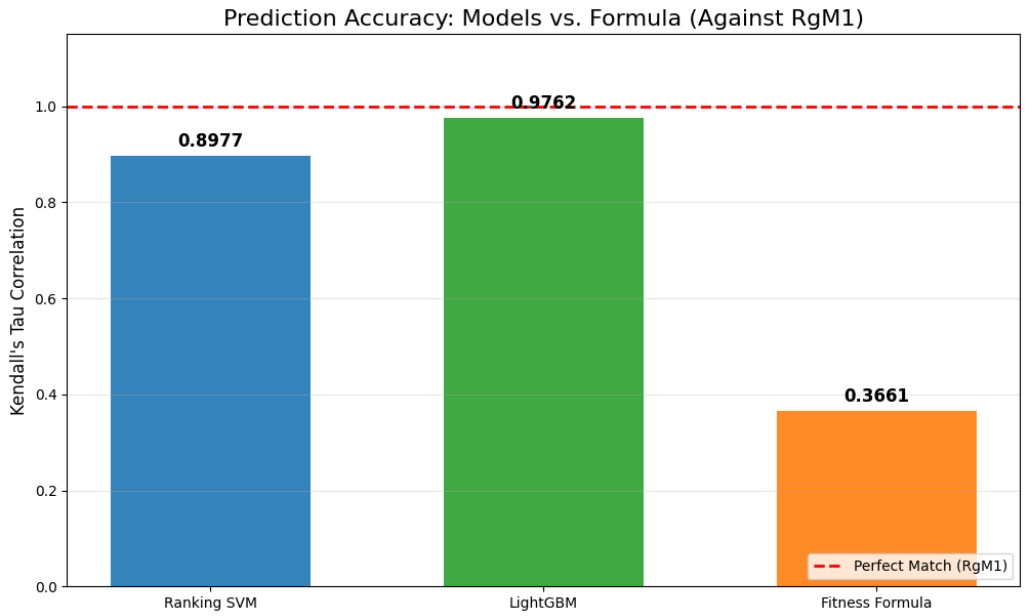

Figure 3.3: Final Comparison of Kendall's Tau Correlation across models. LightGBM outperforms both the Baseline Formula and Ranking SVM.

Conclusion

Our validation process confirms that the LightGBM model is not only the most accurate ($\tau \approx 0.98$) but also remarkably robust ($\tau \approx 0.97$ on new users). This combination of high performance and stability makes it the ideal candidate for a real-time surrogate fitness function in Interactive Evolutionary Computation.

CHAPTER 4

DISCUSSION AND PERSPECTIVES

Plan

4.1 Introduction

In the previous chapter, we presented the experimental results, highlighting the superior performance of the LightGBM model ($\tau \approx 0.97$) compared to the Ranking SVM ($\tau \approx 0.89$) and the Heuristic Baseline ($\tau \approx 0.36$).

In this final chapter, we interpret these findings in the broader context of Interactive Evolutionary Computation (IEC). We specifically address the research question regarding user interface design, discuss the robustness of our validation strategy, and outline the necessary steps to transition this "Offline" feasibility study into a "Online" real-time system.

4.2 Interpretation of Results

The primary objective was to determine if an implicit surrogate model could accurately replace human evaluation. The results provide a decisive validation of this hypothesis.

4.2.1 Non-Linearity of User Preference

The poor performance of the Heuristic Baseline confirms that human gaze behavior is fundamentally non-linear. A user does not simply "look longer at what they like." Complex interactions—such as a short fixation accompanied by a rapid change in pupil diameter—often signal preference more reliably than duration alone. Tree-based models like LightGBM excel at capturing these conditional dependencies, which explains their dominance over the linear baseline and the kernel-based SVM.

4.3 Implications for User Interface Design

A key requirement of this project was to determine the most effective interaction technique for the user: presenting solutions simultaneously (Grid View) or in small groups (Pairwise View). While our experiments focused on machine learning algorithms, the results offer direct evidence to support a specific UI design.

4.3.1 Algorithmic Support for Grid Views

We tested two distinct learning approaches that mirror these UI paradigms:

- **Pairwise Approach (SVM):** This models preference as a series of binary choices ($A > B$). It corresponds to a UI where users compare two images side-by-side.
- **Listwise Approach (LightGBM):** This models preference as a relative ranking of an entire group. It corresponds to a Grid UI where the user scans the whole population.

While the Pairwise SVM performed well ($\tau \approx 0.89$), it was outperformed by the Listwise LightGBM ($\tau \approx 0.97$). The success of the Listwise model proves that the gaze data contains sufficient signal to rank an entire generation at once. **Conclusion:** A **Grid View UI** is scientifically viable. We do not need to restrict the user to tedious 2-by-2 comparisons to obtain accurate data. A Grid View is therefore recommended, as it allows for faster evaluation and reduces physical fatigue (fewer clicks) compared to a Pairwise interface.

4.4 Robustness and Validation

To ensure our results were not due to overfitting, we employed a rigorous two-tier validation strategy specifically designed for evolutionary data.

4.4.1 Temporal Generalization

By strictly separating the Training Set (Past Generations) from the Test Set (Future Generations), we simulated the "Time Arrow" of a real optimization session. The high accuracy on the test set proves that the model can handle **Concept Drift**—it successfully predicts future preferences based solely on past interactions.

4.4.2 Subject Independence

Through our **Leave-Subjects-Out Cross-Validation**, we observed that the model maintains high accuracy even for users it has never seen before. This implies that there are universal gaze patterns (e.g., pupil dilation upon interest) shared across different individuals, making the system scalable to new users without extensive re-calibration.

4.5 Future Work

Based on these findings, we propose the following roadmap.

4.5.1 Real-Time Integration

The immediate next step is to embed the trained LightGBM model into the IEC software loop.

1. **Observation Phase:** The user views the population grid for 5–10 seconds.
2. **Implicit Ranking:** The model predicts the fitness of all individuals instantly.
3. **Evolution:** The algorithm generates the next generation automatically.

Conclusion

We have established that implicit gaze analysis is a powerful and robust tool for evolutionary computation. By validating the Listwise approach, we have provided the evidence needed to design efficient Grid-based interfaces that can theoretically run optimization tasks for longer periods without exhausting the user.

General Conclusion

Interactive Evolutionary Computation (IEC) has long been hampered by the "Human Bottleneck"—the fatigue caused by requiring users to manually evaluate thousands of potential solutions. The objective of this internship was to develop an implicit fitness function capable of predicting user preferences using eye-tracking data, thereby automating the evaluation process.

Summary of Contributions

To address this challenge, we developed a complete machine learning pipeline:

- **Data Analysis:** We extracted and cleaned 21 physiological features (fixations, saccades, pupil diameter) from raw eye-tracking signals.
- **Methodology:** We implemented and compared three distinct approaches: a Heuristic Baseline, Ranking SVM, and LightGBM.
- **Validation:** We designed a rigorous evaluation protocol using Temporal Split and Subject-Group Cross-Validation to ensure the results were realistic and robust.

Key Results

The experiments yielded decisive results. The proposed **LightGBM** model achieved a Kendall's Tau correlation of $\tau \approx 0.97$, vastly outperforming the heuristic baseline ($\tau \approx 0.36$) and Ranking SVM ($\tau \approx 0.89$). This demonstrates that machine learning can accurately decode the complex, non-linear relationship between eye movements and human preference.

Perspectives

These results open the door to a new generation of "Fatigue-Free" evolutionary algorithms. By integrating this surrogate model into the optimization loop, we can theoretically run optimization tasks for much longer periods, exploring deeper solution spaces without exhausting the user.

Future work should focus on the live integration of this model . Ultimately, this work contributes a significant step towards making human-computer optimization more natural, efficient, and seamless.

References

[1] Denis Pallez et al. “Eye-tracking evolutionary algorithm to minimize user’s fatigue in IEC applied to interactive one-max problem”. In: *GECCO ’07: Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation*. ACM. London, United Kingdom, 2007, pp. 2883–2886. DOI: 10.1145/1274000.1274098.

[2] Hideyuki Takagi. “Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation”. In: *Proceedings of the IEEE* 89.9 (2001), pp. 1275–1296.