Evolution-Based Feature Selection for Predicting Dissolved Oxygen Concentrations in Lakes

Runlong Yu[1] , Robert Ladwig[2] , Xiang Xu[3] , Peijun Zhu[4] ,
Paul C. Hanson[5] , Yiqun Xie[6] , and Xiaowei Jia[1(✉)]

[1] University of Pittsburgh, Pittsburgh, USA
{ruy59,xiaowei}@pitt.edu
[2] Aarhus University, Aarhus, Denmark
rladwig@ecos.au.dk
[3] University of Science and Technology of China, Hefei, China
demon@mail.ustc.edu.cn
[4] Georgia Institute of Technology, Atlanta, USA
zpj@gatech.edu
[5] University of Wisconsin-Madison, Madison, USA
pchanson@wisc.edu
[6] University of Maryland, College Park, USA
xie@umd.edu

Abstract. Accurate prediction of dissolved oxygen (DO) concentrations in lakes requires a comprehensive study of phenological patterns across ecosystems, highlighting the need for precise selection of interactions amongst external factors and internal physical-chemical-biological variables. This paper presents the *Multi-population Cognitive Evolutionary Search (MCES)*, a novel evolutionary algorithm for complex feature interaction selection problems. MCES allows models within every population to evolve adaptively, selecting relevant feature interactions for different lake types and tasks. Evaluated on diverse lakes in the Midwestern USA, MCES not only consistently produces accurate predictions with few observed labels but also, through gene maps of models, reveals sophisticated phenological patterns of different lake types, embodying the innovative concept of "AI from nature, for nature".

Keywords: Ecosystem modeling · Adaptive learning · Feature selection

1 Introduction

The concentration of dissolved oxygen (DO) in lakes is a key indicator of water quality and the health of freshwater ecosystems. Effective DO monitoring is critical for sustaining aquatic biodiversity and ensuring water security for human consumption [37]. DO concentrations are influenced not only by the exchange

of oxygen between air and water, but also by the metabolic processes of primary production and respiration [31]. As articulated by Edward A. Birge one century ago [1]: The fluctuations in a lake's oxygen illustrate its "life cycle" more clearly than many other ecological indicators. This is particularly evident in nutrient-rich eutrophic lakes, where algal blooms can significantly deplete oxygen, creating detrimental "dead zones" for aquatic life.

Accurate prediction of DO concentrations requires a comprehensive study of different phenological patterns across various ecosystems. In particular, DO concentration is closely intertwined with ecosystem phenology, influenced by morphometric and geographic information, mass fluxes, weather conditions, trophic state, and watershed land use. In deeper lakes, for instance, light scarcity and decreased mixing with the oxygen-rich surface can lower oxygen [27,30]. Temperature fluctuations impact oxygen solubility and biochemical activities [33]. Land use changes reshape DO patterns and metabolism phenology [11,38].

Given the importance of predicting DO concentration, scientists across limnology, hydrology, meteorology, and environmental engineering have devised process-based models to simulate the dynamics of freshwater ecosystems. These models, aimed at evaluating the effects of external and internal factors, often combine hydrodynamic and water quality models [10]. Examples include DYRESM [7], GLM [8], MyLake [28]. These models utilize first-order principles (e.g., mass and energy conservation), but also involve many parameterizations or approximations due to incomplete physical knowledge or excessive complexity in modeling certain complex processes, resulting in inherent model bias.

Advanced data-driven methods like deep learning [17], offer an alternative to process-based models for complex scientific problems (e.g., prediction of DO concentration). Their success is contingent upon effective feature selection [2,4,18], however, most extant methods face several major challenges in this problem. Firstly, predicting DO involves sophisticated phenological patterns across various metabolic processes. Directly considering a large number of feature interactions can easily introduce noise, while manual selection risks overlooking critical details. Secondly, most feature selection approaches rely on global models built under expert guidance and lack the flexibility to adapt to various tasks and datasets. As a result, they often fail to select relevant feature interactions for different lake types and prediction tasks. Finally, the sparse nature of DO data further complicates model training, as frequent and comprehensive data collection is hindered by substantial human labor and material costs.

To address these challenges, cognitive evolutionary search (CELS) has been developed [43], utilizing an evolutionary algorithm to adaptively evolve models for selecting feature interactions under task-specific guidance. Unlike conventional methods constrained by model fitness evaluation [35,41,45,46], CELS introduces a novel model fitness diagnosis technique. Despite its advancements, CELS still contends with issues of "reproductive isolation", which may result in the convergence of similar models within a population, thereby limiting the development of specialized models for distinct tasks [19,34].

Leveraging insights from CELS, this paper proposes an advanced evolutionary algorithm, namely *Multi-population Cognitive Evolutionary Search (MCES)*, where cognitive ability refers to the malleability of organisms to orientate to the environment [43]. MCES utilizes multi-population models to effectively serve diverse lake types and predictive tasks, mirroring the adaptive strategies of species in diverse habitats. In MCES, feature interactions are envisaged as genomes and models as organisms within ecosystems, conceptualizing tasks as the natural environments these organisms inhabit. Internally, MCES assesses the capacities of genes, with mutations occurring probabilistically when existing traits are detrimental to survival. The models within MCES undergo crossover and mutation within their respective populations and, though less commonly, engage in inter-population crossover. This genetic diversity and flexibility allow models to dynamically adapt and select appropriate phenological feature interactions, catering to specific environmental conditions of different lake types and tasks.

2   Related Work

Existing research suggests that evolution-based feature selection is often limited to filters and wrappers due to model fitness evaluation constraints [35,41]. Specifically, wrapper methods use the performance of the learning algorithm as its evaluation criterion, while filter methods use the intrinsic characteristics of the data. On the other hand, embedded approaches simultaneously select features and learn a classifier, therefore conventional algorithms cannot evaluate the fitness of the model [35,41]. Only genetic programming (GP) and learning classifier systems (LCSs) are able to perform embedded feature selection, but they are not practical [6,20,26,35,41]. For additional research on evolution-based feature selection, please refer to [43].

Research on the population structure of EAs has demonstrated the benefits of segmenting the initial population into multiple sub-populations. These sub-populations exchange information, and regrouping operators are triggered at regular intervals to maintain the population's diversity and balance exploitation with exploration capabilities [39,47]. Various models like the shuffle or update parallel differential evolution (SOUPDE) [36] and the multi-population-based cooperative coevolutionary algorithm (MPCCA) [29] utilize unique mutation strategies and population dynamics to optimize performance across diverse problem-solving scenarios.

3   Problem Definition and Preliminaries

Problem Definition. Our goal is to predict the DO concentration at a daily scale. We simplify our analysis by dividing the water column into two distinct layers with separate oxygen and metabolic kinetics: the epilimnion (upper surface layer) and the hypolimnion (lower bottom layer). We treat the DO prediction for the epilimnion and hypolimnion as two tasks.

For each lake, we have access to its phenological features $\boldsymbol{x}_t$ on each date t. These features, spanning m diverse fields $\boldsymbol{x}_t = \{x_t^1, \cdots, x_t^m\}$, encompass morphometric and geographic details such as lake area, depth, and shape; flux-related data like ecosystem and sedimentation fluxes; weather factors comprising wind speed and temperature; a range of trophic states from dystrophic to eutrophic; and diverse land use proportions extending from forests to wetlands. Besides input features, we also have observed DO concentrations y_t (for both the epilimnion y_t^{epi} and hypolimnion y_t^{hyp}) on certain days.

We use an embedding layer to convert input phenological features into a series of multi-field feature embeddings $\boldsymbol{f}_t = [\boldsymbol{f}_t^1, \cdots, \boldsymbol{f}_t^m]$, where $\boldsymbol{f}_t^i = \mathrm{embed}(x_t^i)$. Initially, numerical features are bucketed into categories, and each category is then represented as a one-hot vector that is transformed into an embedding vector through a perceptron layer [44]. Our model uses these embeddings to predict DO concentrations $\hat{y}_t$ for both the epilimnion $\hat{y}_t^{\mathrm{epi}}$ and hypolimnion $\hat{y}_t^{\mathrm{hyp}}$.

Feature Interaction Selection. The process of feature interaction selection aims to identify the most informative feature interactions that can facilitate the prediction of target DO concentrations $\mathcal{H} : \mathcal{M}(\boldsymbol{f}, \boldsymbol{g}(\boldsymbol{f})) \rightarrow \hat{y}$, where $\boldsymbol{g}$ denotes the set of operations to interact on feature pairs, and $\boldsymbol{g}(\boldsymbol{f})$ denotes the set of interactions. In DO prediction case, the algorithm $\mathcal{H}$ aims to minimize the MSE loss for the outputs of the prediction model $\mathcal{M}$, given as:

$$\mathcal{L}(\mathcal{M}) = \frac{1}{|B|} \sum_{t \in B} \left(y_t - \hat{y}_t \right)^2, \tag{1}$$

where B denotes the set of instance indices in a mini-batch, $\hat{y}_t$ denotes the predictive result given through the learned model.

As the fundamental components in feature interaction, operations are functions where two individual features are converted into an interaction. For the sake of simplicity, we adopt four representative operations as candidate operations to present instantiations of MCES, i.e., $\boldsymbol{g} = \{\oplus, \otimes, \boxplus, \boxtimes\}$. Detailed explanations of these operations and their use in evolution-based feature selection are provided in **Appendix A**.

Metabolic Process-Based Model. In this work, we use a metabolic process-based model to simulate DO labels [16]. The process-based model divides the water column into the upper epilimnion and lower hypolimnion during stratified periods. It is focused on analyzing metabolic dynamics in warmer months. Flux features (denoted as F) affecting DO concentrations are calibrated with observed data [16]. More details of the process-based model are provided in **Appendix B**.

4   Overall Framework

The overall framework is depicted in Fig. 1 and involves two stages of learning: (1) MCES for feature interaction selection using simulated labels, and (2) model refinement using real observed labels.
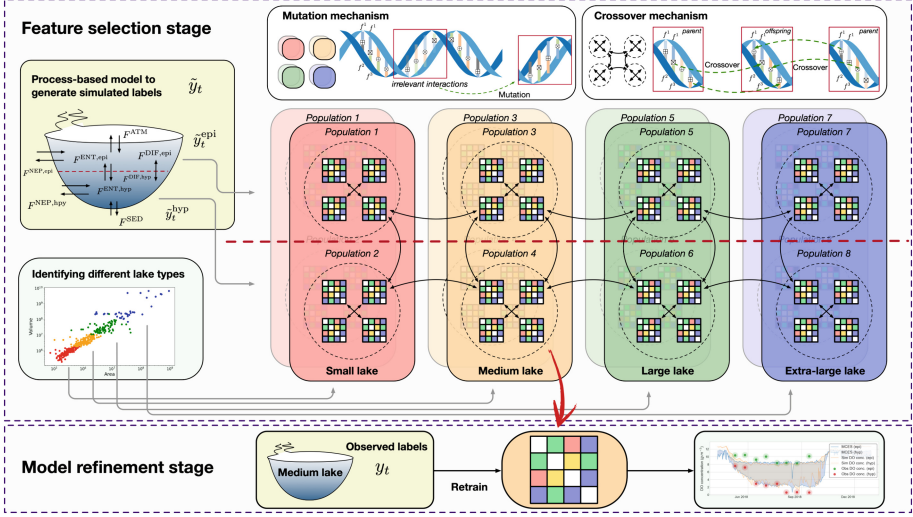
Fig. 1. Overall framework.

Specifically, we first utilize metabolic process-based models to generate simulated DO labels given the phenological features $\boldsymbol{x}_t$. Using these simulated labels, we implement our proposed feature interaction selection algorithm MCES. The models in MCES adaptively evolve to select relevant feature interactions within populations for different lake types and tasks. In the subsequent stage, we refine these evolved models using real-world observed DO concentration data. This mirrors natural genetic decoding, where selected feature interactions are further optimized to reflect actual ecological dynamics.

Identifying Different Lake Types. As a preprocessing step, lakes are classified based on characteristics critical to oxygen dynamics-primarily surface area and volume. Larger surface areas improve atmospheric oxygen exchange, while greater volumes are associated with higher oxygen consumption in deeper waters. We use a balanced K-means clustering algorithm to uniformly distribute the lakes in dataset $\boldsymbol{L}$ into four categories based on volume and area: small lakes $\boldsymbol{L}_{\mathrm{S}}$, medium lakes $\boldsymbol{L}_{\mathrm{M}}$, large lakes $\boldsymbol{L}_{\mathrm{L}}$, and extra-large lakes $\boldsymbol{L}_{\mathrm{xL}}$.

5   Multi-Population Cognitive Evolutionary Search

In this section, we detail the feature selection stage, employing simulated labels $\tilde{y}_t$ to implement MCES. We begin by introducing the establishment of multiple populations, and explain how each individual within these populations is modeled. We then delve into the fundamental mutation and crossover mechanisms. Lastly, we provide an instantiation of the search process.

5.1   Multiple Populations

In MCES, each population comprises a set of models designed specifically for a particular combination of lake type and prediction task (e.g., predicting DO in the epilimnion of large lakes). Consequently, we establish eight unique populations: $\boldsymbol{P}_\mathrm{S}^\mathrm{epi}$, $\boldsymbol{P}_\mathrm{S}^\mathrm{hyp}$, $\boldsymbol{P}_\mathrm{M}^\mathrm{epi}$, $\boldsymbol{P}_\mathrm{M}^\mathrm{hyp}$, $\boldsymbol{P}_\mathrm{L}^\mathrm{epi}$, $\boldsymbol{P}_\mathrm{L}^\mathrm{hyp}$, $\boldsymbol{P}_\mathrm{xL}^\mathrm{epi}$, $\boldsymbol{P}_\mathrm{xL}^\mathrm{hyp}$. Here the subscript $\{\mathrm{S},\mathrm{M},\mathrm{L},\mathrm{xL}\}$ represents different lake types, and the superscript $\{\mathrm{epi},\mathrm{hyp}\}$ represents two different prediction tasks (i.e., epilimnion or hypolimnion).

   With a population size of n (where $n > 1$), we initialize each population with n models for feature interaction selection: $\boldsymbol{P} = \{\mathcal{M}_1, \cdots, \mathcal{M}_n\}$. We conceptualize each model as a natural organism that evolves to enhance its traits for improved fitness within its environment. These traits, inherited through the organism's genomes, stem from the interplay between features and operations, much like nucleotides and their connections. Following various linkages of nucleotides, we extend the operation set with four types of operations as the search space, i.e., $\boldsymbol{g} = \{\oplus, \otimes, \boxplus, \boxtimes\}$. If g_k is a chosen operation from $\boldsymbol{g}$, an interaction $g_k(\boldsymbol{f}_t^i, \boldsymbol{f}_t^j)$ is defined by applying the operation g_k to a pair of features $(\boldsymbol{f}_t^i, \boldsymbol{f}_t^j)$.

   Motivated by the goal of enhancing model fitness through the preservation of beneficial genetic information, we aim to discern and prioritize important features and their interactions via a parameterized method, called internal genetic evaluation. The idea is to introduce a set of relevance parameters to strengthen relevant feature interactions while diminishing or mutating those that contribute less. In this context, we define relevance parameters for features $\boldsymbol{f}_t$ and interactions $\tilde{g}(\boldsymbol{f}_t)$ as $\boldsymbol{\alpha} = \{\alpha_i | 1 \leqslant i \leqslant m\}$ and $\boldsymbol{\beta} = \{\beta_{i,j} | 1 \leqslant i < j \leqslant m\}$, respectively. Here, $\tilde{g}(\boldsymbol{f}_t)$ denotes the interaction of applying any operations from $\boldsymbol{g}$ to a pair of features. The predictive response of our model at time step t is formulated as:

$$\hat{y}_t = \mathcal{M}\big(\boldsymbol{\alpha} \cdot \boldsymbol{f}_t, \boldsymbol{\beta} \cdot \tilde{g}(\boldsymbol{f}_t)\big), \tag{2}$$

where $\mathcal{M}$ can be any individual model in the population. In this work, we use a sequence encoder with Long-Short Term Memory (LSTM) networks [9] to efficiently encode temporal information and the dynamics of feature interactions. The model $\mathcal{M}$ is thus depicted as:

$$\begin{aligned} \boldsymbol{h}_t^\iota &= \mathrm{LSTM}\Big([\boldsymbol{\alpha} \cdot \boldsymbol{f}_t, \boldsymbol{\beta} \cdot \tilde{g}(\boldsymbol{f}_t)]; \boldsymbol{h}_{t-1}^\iota\Big), \\ \hat{y}_t &= \boldsymbol{W}^\iota \cdot \boldsymbol{h}_t^\iota + \boldsymbol{b}^\iota, \end{aligned} \tag{3}$$

where $\boldsymbol{h}_t^\iota$ represents a series of hidden states, and $\boldsymbol{W}^\iota$ and $\boldsymbol{b}^\iota$ denote the weight and bias parameters, respectively. The loss function for model $\mathcal{M}$, calculated using simulated labels $\tilde{y}_t$, is defined as:

$$\mathcal{L}(\mathcal{M}) = \frac{1}{|B|} \sum_{t \in B} \Big(\tilde{y}_t - \hat{y}_t\Big)^2, \tag{4}$$

where B denotes the set of instance indices within a mini-batch.

We use a regularized dual averaging (RDA) optimizer to learn the relevance parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with the aim to distinguish between relevant and irrelevant feature interactions through this process [3,40]. When the absolute value of the cumulative gradient average value in a certain position in $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ is less than a threshold, the weight of that position in relevance parameters will be set to 0, resulting in the sparsity of the relevance [22,40]. Meanwhile, the parameters of the embedding layer (for $\boldsymbol{f}_t$) are optimized using the Adam optimizer [14]. Unlike AutoML [23,24], which categorizes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as high-level decisions and treats feature embeddings as lower-level variables for bi-level optimization, our approach simplifies this process. To circumvent the complex and costly bi-level optimization, we update feature embeddings and relevance parameters jointly using single-level optimization with gradients on the training set, as $\nabla_{\boldsymbol{f}}\mathcal{L}(\boldsymbol{f}_{iter-1},\boldsymbol{\alpha}_{iter-1},\boldsymbol{\beta}_{iter-1})$ and $\nabla_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathcal{L}(\boldsymbol{f}_{iter-1},\boldsymbol{\alpha}_{iter-1},\boldsymbol{\beta}_{iter-1})$, respectively.

5.2   Mutation Mechanism and Crossover Mechanism

With our definitions of population and feature interaction selection models, we further detail the mutation and crossover mechanisms in MCES. The crossover mechanism is bifurcated into intra-population and inter-population crossover.

Mutation Mechanism. The mutation serves as a fundamental mechanism of our search process. It primarily aims at mutating the operations associated with irrelevant interactions into alternative operations, and thus generating a new model (the offspring). Specifically, for an interaction $g_k(f_t^i, f_t^j)$, mutation is triggered with a probability σ after every τ steps if the relevance parameter $\beta_{i,j}$ drops below a threshold λ. In other words, to regenerate a new interaction, the operation g_k of the interaction $g_k(f_t^i, f_t^j)$ mutates into another operation g_l, given as:

$$g_k = \begin{cases} g_l \text{ with probability } \sigma, & \text{if } \beta_{i,j} < \lambda, \\ g_k, & \text{otherwise,} \end{cases} \tag{5}$$

where g_l is randomly selected from the operation set as $g_l = \{g \mid g \in \boldsymbol{g}, g \neq g_k\}$. The new interaction $g_l(f_i, f_j)$ replaces the irrelevant interaction $g_k(f_i, f_j)$, and its corresponding relevance $\beta_{i,j}$ is reset. Consequently, the parent model $\mathcal{M}$ evolves into its offspring $\mathcal{M}'$, which incorporates these fresh interactions with revised relevance $\boldsymbol{\beta}'$, and maintains features with relevance $\boldsymbol{\alpha}'$ inherited from $\boldsymbol{\alpha}$.

Intra-Population Crossover Mechanism. Given a population $\boldsymbol{P} = \{\mathcal{M}_1, \cdots, \mathcal{M}_\nu, \cdots, \mathcal{M}_n\}$, we use $\boldsymbol{\beta}^{\mathcal{M}_\nu}$ to denote the relevance parameters of interactions for each model $\mathcal{M}_\nu$. The obtained $\boldsymbol{\beta}^{\mathcal{M}_\nu}$ can vary across different models in $\boldsymbol{P}$. Therefore, within this population, the models may have a variety of operations for interacting with each feature pair (f_i, f_j), represented as $g_{i,j}^{\boldsymbol{P}} = \{g_{i,j}^{\mathcal{M}_1}, \cdots, g_{i,j}^{\mathcal{M}_\nu}, \cdots, g_{i,j}^{\mathcal{M}_n}\}$. The intra-population crossover mechanism aims to select the most suitable operation (of which interaction has the largest

relevance) within the population to apply on the feature pair for the offspring model $\mathcal{M}'$, given as:

$$g_{i,j}^{\mathcal{M}'} = \arg \max_{g_{i,j}^{\mathcal{M}_\nu} \in g_{i,j}^{\boldsymbol{P}}} \beta_{i,j}^{\mathcal{M}_\nu}. \tag{6}$$

Meanwhile, the relevance parameters of interactions in this offspring model are inherited from their respective parent models (i.e., the selected $\mathcal{M}_\nu$).

Inter-Population Crossover Mechanism. For two distinct populations $\boldsymbol{P}_A = \{\mathcal{M}_1^A, \cdots, \mathcal{M}_\nu^A, \cdots, \mathcal{M}_n^A\}$ and $\boldsymbol{P}_B = \{\mathcal{M}_1^B, \cdots, \mathcal{M}_\nu^B, \cdots, \mathcal{M}_n^B\}$, we use $\boldsymbol{\beta}^{\mathcal{M}_\nu^A}$ and $\boldsymbol{\beta}^{\mathcal{M}_\nu^B}$ to denote the relevance of interactions for two populations $\boldsymbol{P}_A$ and $\boldsymbol{P}_B$, respectively. The inter-population crossover mechanism works as follows: For each feature pair (f_i, f_j), we select the most suitable operation from $\boldsymbol{P}_B$ to interact on the feature pair in the offspring model of $\boldsymbol{P}_A$. Conversely, the most suitable operation of the feature pair from $\boldsymbol{P}_A$ is selected for the offspring model of $\boldsymbol{P}_B$, given as:

$$g_{i,j}^{\mathcal{M}'_A} = \arg \max_{g_{i,j}^{\mathcal{M}_\nu^B} \in g_{i,j}^{\boldsymbol{P}_B}} \beta_{i,j}^{\mathcal{M}_\nu^B}, \quad g_{i,j}^{\mathcal{M}'_B} = \arg \max_{g_{i,j}^{\mathcal{M}_\nu^A} \in g_{i,j}^{\boldsymbol{P}_A}} \beta_{i,j}^{\mathcal{M}_\nu^A}. \tag{7}$$

Meanwhile, the relevance parameters of interactions in the offspring models are inherited from their respective parent models.

5.3 Instantiation of the Search Process

Utilizing mutation, intra- and inter-population crossover mechanisms, we implement MCES as detailed in Algorithm 1. MCES begins by randomly initializing eight distinct model populations (line 1). It then follows a series of iterative steps (lines 6–28), continuing until convergence. Each iteration involves optimizing offspring models and their relevance parameters within each population.

For every τ iterations, the algorithm (lines 9–16) selects and replaces the worst model $\mathcal{M}$ in each population $\boldsymbol{P}$ based on the designated loss function (Eq. (4)). The selection process can be expressed as:

$$\mathcal{M} = \arg \max_{\mathcal{M}_\nu \in \boldsymbol{P}} \mathcal{L}(\mathcal{M}_\nu). \tag{8}$$

When the algorithm replaces the worst model $\mathcal{M}$ with the offspring model $\mathcal{M}'$, a new offspring $\mathcal{M}'$ is generated through intra-population crossover and subsequent mutation, enhancing genotypic diversity, thus enabling the search process to effectively avoid local optima and explore global regions.

For every $ep \times \tau$ iterations (lines 18–27), we randomly select a pair of populations, $\boldsymbol{P}_A$ and $\boldsymbol{P}_B$, either based on a shared task (i.e., epilimnion or hypolimnion) across different lake types or on the same lake type but with different tasks, with selection probability balanced by a coin flip. This leads to the generation of new offspring models $\mathcal{M}'_A, \mathcal{M}'_B$ through inter-population crossover between $\boldsymbol{P}_A, \boldsymbol{P}_B$, promoting the exchange of advantageous genotypic patterns across different lake

---

Algorithm 1. Multi-population Cognitive Evolutionary Search

Input: Training dataset of four types of lakes $\boldsymbol{L}_{\mathrm{S}}$, $\boldsymbol{L}_{\mathrm{M}}$, $\boldsymbol{L}_{\mathrm{L}}$, $\boldsymbol{L}_{\mathrm{xL}}$, each lake has features $\boldsymbol{f}_t$, simulated DO labels $\tilde{y}_t^{\mathrm{epi}}$, $\tilde{y}_t^{\mathrm{hyp}}$ over T days; operation set $\boldsymbol{g}$.

1: Initialize eight populations $\boldsymbol{P}_{\mathrm{S}}^{\mathrm{epi}}$, $\boldsymbol{P}_{\mathrm{S}}^{\mathrm{hyp}}$, $\boldsymbol{P}_{\mathrm{M}}^{\mathrm{epi}}$, $\boldsymbol{P}_{\mathrm{M}}^{\mathrm{hyp}}$, $\boldsymbol{P}_{\mathrm{L}}^{\mathrm{epi}}$, $\boldsymbol{P}_{\mathrm{L}}^{\mathrm{hyp}}$, $\boldsymbol{P}_{\mathrm{xL}}^{\mathrm{epi}}$, $\boldsymbol{P}_{\mathrm{xL}}^{\mathrm{hyp}}$, of which any $\mathcal{M}$ has initialized $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
2: **for** each $\boldsymbol{P}$ **do**
3:     Generate $\mathcal{M}'$ via intra-population crossover in $\boldsymbol{P}$.   ▷ Eq. (6)
4:     Mutate $\mathcal{M}'$.   ▷ Eq. (5)
5: **end for**
6: **while** not converged **do**
7:     **for** each $\boldsymbol{P}$ **do**
8:         Optimize $\mathcal{M}'$ with $\boldsymbol{\alpha}'$ $\boldsymbol{\beta}'$.
9:         **if** $\mathrm{mod}(t, \tau) = 0$ **then**
10:             Select the worst $\mathcal{M}$.   ▷ Eq. (8)
11:             Replace $\mathcal{M}$ in $\boldsymbol{P}$ with $\mathcal{M}'$.
12:             **if** $\mathrm{mod}(t, ep \times \tau) \neq 0$ **then**
13:                 Generate $\mathcal{M}'$ via intra-population crossover.
14:                 Mutate $\mathcal{M}'$.   ▷ Eq. (5)
15:             **end if**
16:         **end if**
17:     **end for**
18:     **if** $\mathrm{mod}(t, ep \times \tau) = 0$ **then**
19:         Choose $(\boldsymbol{P}_A, \boldsymbol{P}_B)$ either by task or lake type.
20:         Generate $(\mathcal{M}'_A, \mathcal{M}'_B)$ via inter-population crossover of $\boldsymbol{P}_A$, $\boldsymbol{P}_B$.   ▷ Eq. (7)
21:         **for** each $\boldsymbol{P}$ not in $(\boldsymbol{P}_A, \boldsymbol{P}_B)$ **do**
22:             Generate $\mathcal{M}'$ via intra-population crossover.
23:         **end for**
24:         **for** each $\boldsymbol{P}$ **do**
25:             Mutate $\mathcal{M}'$.   ▷ Eq. (5)
26:         **end for**
27:     **end if**
28: **end while**
29: **return** the set of best models $\mathcal{M} = \arg\min_{\mathcal{M}_\nu \in \boldsymbol{P}} \mathcal{L}(\mathcal{M}_\nu)$ from each population $\boldsymbol{P}$.

---

types and tasks. Meanwhile, each remaining population generates its offspring, $\mathcal{M}'$, through intra-population crossover, followed by mutation of all offspring.

Finally, the algorithm concludes by delivering a set of the best models, one from each population (line 29), thereby ensuring a comprehensive exploration and exploitation of the search space across diverse environmental contexts.

6   Model Refinement

Inspired by nature's replication and transcription processes, which translate genetic information into protein sequences to equip organisms with diverse functions, we proceed to a model refinement stage. Here our objective is to refine the model to better leverage the features and interactions obtained from MCES. At

this stage, we select the corresponding model by the lake type and task, and then use observed labels for the model refinement. Relevant features and interactions are selected according to their relevance parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$. If $\alpha_i = 0$ or $\beta_{i,j} = 0$, the corresponding features or interactions are fixed to be discarded permanently. Given the scarcity of observed data, we inherit parameters from the preceding LSTM to ensure the model's effective learning, given as:

$$
\begin{aligned}
\boldsymbol{h}_t^o &= \text{LSTM}\Big([\boldsymbol{\alpha} \cdot \boldsymbol{f}_t, \boldsymbol{\beta} \cdot \tilde{g}(\boldsymbol{f}_t)]; \boldsymbol{h}_{t-1}^o\Big) \\
\hat{y}_t &= \boldsymbol{W}^o \cdot \boldsymbol{h}_t^o + \boldsymbol{b}^o
\end{aligned}
\tag{9}
$$

where $\boldsymbol{h}_t^o$ represents a series of hidden states, with $\boldsymbol{W}^o$ and $\boldsymbol{b}^o$ as the weight and bias parameters. The relevance $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are fixed and serve as attention units.

To address the disparity between abundant simulated and scarce observed labels, we've developed a new loss function for LSTM that integrates both types of data through weighted imputation [12,42]. This assigns a greater weight to the loss on observed data and a smaller weight to simulated data, effectively addressing the scarcity of observed labels. The loss function is expressed as:

$$
\mathcal{L}(\mathcal{M}) = \frac{1}{|B|} \sum_{t \in B} \mathbb{I}(y_t)\Big(y_t - \hat{y}_t\Big)^2 + \rho\Big(1 - \mathbb{I}(y_t)\Big)\Big(\tilde{y}_t - \hat{y}_t\Big)^2,
\tag{10}
$$

where $\hat{y}_t$ denotes the predicted DO concentration, y_t is the observed DO concentration, $\tilde{y}_t$ is the simulated DO concentration, $\mathbb{I}(x)$ is an indicator function that equals 1 if x is observed (true) and 0 otherwise (false), and ρ is the tradeoff parameter balancing observed and simulated labels.

7   Experimental Evaluation

7.1   Dataset

We evaluate the proposed MCES for predicting DO concentration using a dataset that documents over 41 years of ecological observations from 375 lakes in the Midwestern USA, starting in 1979. This dataset includes around 5.58 million daily records, each with 39 fields of phenological features such as morphometric, flux data, weather conditions, trophic states, and land use details. Observed DO data were sourced from the Water Quality Portal (WQP). Lake residence time was taken from the HydroLAKES. Trophic state probabilities (eutrophic, oligotrophic, dystrophic) were from the dataset [25]. Land use proportions of each lake's watershed were taken from the National Land Cover Database (NLCD). An account of these features is available in **Appendix C**. For training MCES, we use data from all 375 lakes. We then selectively conduct testing on lakes that have the most comprehensive DO observations for each type. For large and extra-large lakes, we use data up to 2017 for training, 2018 for validation, and 2019 for testing. For small and medium lakes, where DO observations in 2019 are relatively sparse, we use data up to 2016 for training, 2017 for validation, and 2018 for testing.

7.2   Baselines

We compare to a set of baselines in our experiment: *Sim DO Conc.*: This baseline is the metabolic process-based model used in our first stage, leveraging few observed labels to calibrate simulations that can significantly augment the data for other baselines. *LSTM*: As adopted in our model refinement stage, LSTM incorporates simulated labels for weighted imputation and backward gradient adjustments, a necessity for convergence given the scarcity of observed labels. *EA-LSTM & KGSSL* [5,15]: These time series prediction models, which assimilate hydrological behavior and physical processes, respectively, are regarded as cutting-edge within hydrological and ecological domains. LSTM, EA-LSTM, and KGSSL use individual features for input without feature interaction modeling. *AutoInt, AutoGroup, & AutoFeature* [13,21,32]: These methods are at the forefront of feature interaction modeling. They have demonstrated their utility and versatility through extensive commercial deployment, showcasing their capacity to model complex feature combinations.

Table 1. Comparative performance of DO concentration (g/m^3) prediction in terms of root mean square error (RMSE) across different lake types and tasks. The mean and standard deviation (displayed in grey) of RMSE are calculated from five runs.

Algo. Name	Small lakes		Medium lakes		Large lakes		Extra-large lakes	
	Epi.	Hyp.	Epi.	Hyp.	Epi.	Hyp.	Epi.	Hyp.
Sim DO conc.	1.943	2.212	1.940	2.217	2.620	2.937	1.536	2.772
	± 0.000	± 0.000	± 0.000	± 0.000	± 0.000	± 0.000	± 0.000	± 0.000
LSTM	1.802	1.973	1.744	2.001	2.298	2.630	1.479	2.594
	± 0.079	± 0.064	± 0.092	± 0.081	± 0.088	± 0.043	± 0.068	± 0.056
EA-LSTM	1.716	1.783	1.676	1.546	2.111	2.629	1.478	2.278
	± 0.047	± 0.098	± 0.084	± 0.054	± 0.045	± 0.043	± 0.039	± 0.062
KGSSL	1.793	1.467	1.557	1.632	2.064	2.730	1.294	2.425
	± 0.044	± 0.057	± 0.062	± 0.094	± 0.103	± 0.060	± 0.047	± 0.075
AutoInt	1.510	1.406	1.516	1.626	1.716	1.924	1.112	1.847
	± 0.080	± 0.097	± 0.088	± 0.094	± 0.072	± 0.078	± 0.081	± 0.085
AutoGroup	1.473	1.509	1.364	1.875	1.384	1.600	0.937	1.953
	± 0.078	± 0.080	± 0.059	± 0.072	± 0.075	± 0.068	± 0.085	± 0.076
AutoFeature	1.382	1.768	1.422	1.467	1.405	1.465	1.178	1.976
	± 0.063	± 0.089	± 0.070	± 0.081	± 0.084	± 0.082	± 0.078	± 0.093
MCES (-refine)	1.851	2.044	1.812	2.024	2.374	2.775	1.530	2.648
	± 0.204	± 0.210	± 0.227	± 0.223	± 0.215	± 0.218	± 0.208	± 0.243
MCES (-multi)	1.315	1.624	1.390	1.473	1.390	1.572	1.151	1.848
	± 0.182	± 0.207	± 0.203	± 0.204	± 0.208	± 0.194	± 0.197	± 0.235
MCES (-inter)	1.107	1.375	1.161	1.354	1.004	1.307	1.040	1.536
	± 0.179	± 0.193	± 0.192	± 0.190	± 0.198	± 0.189	± 0.194	± 0.223
MCES	1.076	1.316	1.060	1.288	0.988	1.243	0.918	1.415
	± 0.146	± 0.161	± 0.137	± 0.159	± 0.169	± 0.156	± 0.171	± 0.215

7.3   Implementation Details

To implement MCES, we perform a grid search setting feature embedding size $|\boldsymbol{f}_t^i| = 15$. We use RDA optimizer [3,40] to discriminate the relevant and irrelevant feature interactions, with the learning rate $\gamma = 10^{-3}$, adjustable hyperparameters $c = 0.5, \mu = 0.8$. We set the population size as $n = 4$. We set the mutation mechanism as the mutation threshold $\lambda = 0.2$, the mutation probability $\sigma = 0.5$, and the mutation step size $\tau = 10$. We set the inter-population crossover step size $ep = 10$, and the tradeoff parameter $\rho = 0.1$. To assess the effectiveness of utilizing multiple populations and refining models, we conduct ablation studies with the following variants: *MCES (-refine)*: This variant drops the model refinement stage, using only the simulated labels from the first stage for training. *MCES (-multi)*: Instead of using multiple populations, this variant trains a single population on all data, and tests it separately across different lake types. This setup also excludes the inter-population crossover mechanism. *MCES (-inter)*: This variant does not include the inter-population crossover mechanism.
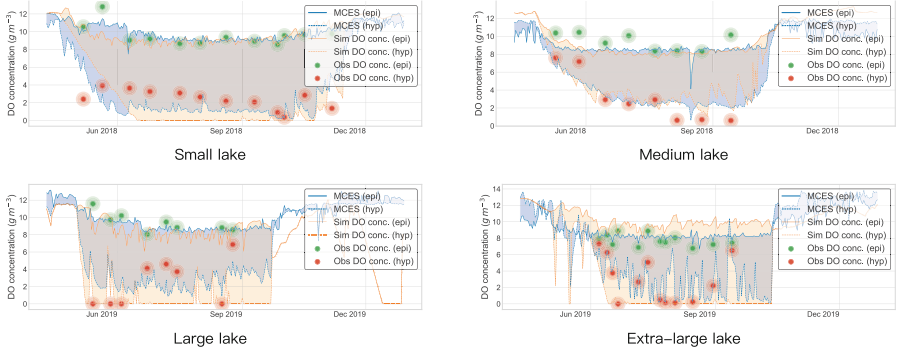

Fig. 2. Time-series analysis of DO concentrations: a comparison of predicted (MCES), simulated, and observed values.

7.4   Experimental Results

Performance Comparison. Table 1 presents a comparative analysis of MCES against various baselines, utilizing root mean square error (RMSE) across diverse lake types and tasks, with both mean and standard deviation calculated over five runs. From the results, we have the following key observations: First, machine learning models universally outperform simulations alone, underscoring the value of integrating observed labels with simulated labels for enhanced prediction accuracy. Second, EA-LSTM and KGSSL surpass LSTM in performance, evidencing the advantage of incorporating hydrological behaviors and physical processes into models, particularly when faced with a scarcity of labels. Third, AutoInt,

AutoGroup, and AutoFeature demonstrate the predictive power of feature inter-
actions, offering significant improvements over models that rely solely on indi-
vidual feature inputs. Fourth, MCES outperforms all baseline models, attribut-
ing its success to the adaptive modeling of interactions through evolutionary
operation selection. Unlike other methods that indiscriminately handle all fea-
ture interactions, MCES discerns relevant feature interactions for specific lake
types and tasks, optimizing their impact while minimizing less pertinent ones.
Lastly, ablation study results reveal that MCES performs better than MCES
(-refine), highlighting the value of combining real with simulated labels for accu-
racy. MCES also exceeds MCES (-multi), emphasizing the necessity of distinct
populations for different ecological settings. Moreover, MCES outshines MCES
(-inter), demonstrating the benefits of inter-population interactions. These inter-
actions promote the sharing of effective feature interactions among populations,
thereby improving the algorithm's accuracy and stability.

Figure 2 offers a time-series comparison of predicted (i.e., MCES), simulated,
and observed DO concentrations, with a specific emphasis on the summer season
of the testing period. The analysis reveals that MCES predictions not only align
closely with observed values but also demonstrate sensitivity to subtle features
and interactions, enhancing their accuracy. While simulated DO concentrations
also generally exhibit a clear trend, there are instances of slight deviation from
observed data. Encouragingly, both predicted and simulated values largely mir-
ror observed trends, highlighting the efficacy and significance of our proposed
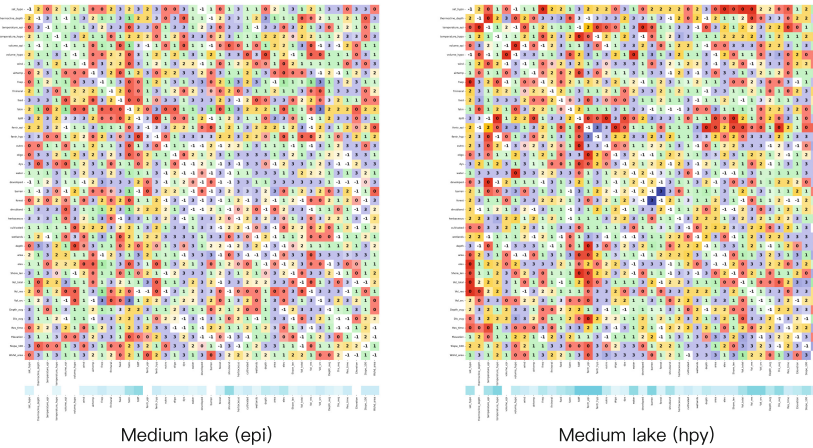MCES in advancing research in this domain.

Medium lake (epi)   Medium lake (hpy)

Fig. 3. Visualization of gene maps for medium lakes.

Visualization of Gene Maps. The model populations accommodate differ-
ent lake types and tasks, leading to a rich diversity in model traits. These

traits influence their survival and fitness rates, mirroring the selection process for operations and feature interactions. To demonstrate the model's evolutionary process and how feature interactions adapt across different lake types and tasks, we visualize the model's **gene maps**. Adopting an encoding where $\oplus = 0, \otimes = 1, \boxplus = 2, \boxtimes = 3$, we can diagnose the model's fitness as a symmetric matrix. Distinct colors are allocated to each operation, creating a vibrant gene map where each gene symbolizes an interaction; like red "0", green "1", yellow "2", and blue "3". For example, a green "1" within the "`depth × area`" block signifies that the element-wise product $\otimes$ is identified as the optimal operation for "`depth`" to interact with "`area`". The intensity of the colors on the gene map is directly correlated with the relevance of the interactions, with darker hues denoting higher relevance and lighter ones suggesting lesser importance. Individual features are also visually encoded as single-hued bars. Interactions deemed irrelevant, with their relevance parameters reduced to 0, are excluded, leaving their corresponding genes depicted in white "−1".

In **Appendix D**, we present gene maps for all lake types and tasks, using end-of-training data to highlight relevant feature interactions for DO concentration prediction. For instance, in Fig. 3, we showcase gene maps for medium lakes.

These maps reveal that, in larger lakes, the DO dynamics are predominantly influenced by sediment oxygen demand and atmospheric exchange, reflecting their extensive water volumes. Conversely, smaller lakes exhibit DO concentrations that are notably impacted by local land use and meteorological factors due to their shallower depths and greater vulnerability to changes in their external watershed environments. Across the board, temperature-related interactions are significant, affecting DO solubility and the lake's biological processes. Additionally, wind speed and atmospheric exchange flux stand out as key drivers of surface gas exchange influencing epilimnion, while the trophic state markers provide indicators of possible oxygen production in the epilimnion and eventual hypolimnetic depletion due to the formation of algal blooms. These findings suggest that diverse ecological factors interplay differently across lake environments, necessitating adaptable prediction models that can cater to these variances.

Additionally, we have included animations in the **supplementary** materials to illustrate the evolution path of MCES. These animations display yearly changes in relevant feature interactions for DO concentrations from 1979. The enduring patterns of feature interactions hint at consistent ecological processes, while deviations in their relevance suggest adaptation to environmental shifts and human activities. Changes in the importance of certain interactions may stem from better land management or climate variations affecting lake stratification. Meanwhile, the emergence of new significant interactions could be a reaction to changes in lake usage or watershed practices. These temporal dynamics underscore the adaptability of MCES, which recalibrates the significance of feature interactions to align with the changing lake environments over time.

In **Appendix E**, we also demonstrate the impact of feature interactions identified by MCES. By adjusting the RDA optimizer's parameters, we consistently

choose a sparser set of feature interactions, accepting a trade-off in accuracy. Simultaneously, a random strategy is applied for comparative purposes, where operations for feature interactions are allocated at random, and some interactions are arbitrarily removed as sparsity intensifies. The gene map showcased at a feature interaction sparsity level around 0.5 offers insight into the model's structure under reduced complexity. This experiment highlights MCES's superior performance even as many feature interactions are discarded, emphasizing its precision in identifying relevant interactions under task guidance. Conversely, the random approach shows a quicker performance drop due to the loss of important interactions. When feature interactions become exceedingly sparse, both methodologies suffer in performance, indicating that a limited set of feature interactions fails to significantly contribute to the model's predictive capabilities. In such cases, individual features primarily drive performance.

8   Conclusion

This paper presents a novel evolutionary algorithm, namely *Multi-population Cognitive Evolutionary Search (MCES)*, blending adaptive learning with natural processes for predicting DO concentrations. MCES employs multi-population models customized for varied lake types and tasks, reflecting the diverse survival strategies of species across various habitats. Evaluated on a variety of lakes in the Midwestern USA, MCES not only demonstrates accurate DO concentration predictions with limited observed data but also reveals sophisticated phenological patterns, highlighting its utility for environmental science and freshwater management. MCES introduces an innovative concept of "AI from nature, for nature". We encourage further exploration and development of new algorithms inspired by MCES, aimed at benefiting the environment.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Birge, E.A.: Gases dissolved in the waters of Wisconsin lakes. Trans. Am. Fish. Soc. **35**(1), 143–163 (1906)
2. Brookhouse, J., Freitas, A.: Fair feature selection with a lexicographic multi-objective genetic algorithm. In: Rudolph, G., Kononova, A., Aguirre, H., Kerschke, P., Ochoa, G., Tušar, T. (eds.) Parallel Problem Solving from Nature – PPSN XVII: 17th International Conference, PPSN 2022, Dortmund, Germany, September 10–14, 2022, Proceedings, Part II, pp. 151–163. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-14721-0_11
3. Chao, S.K., Cheng, G.: A generalization of regularized dual averaging and its dynamics. arXiv preprint arXiv:1909.10072 (2019)

4. Correia, J., Machado, P., Romero, J., Carballal, A.: Feature Selection and Novelty in Computational Aesthetics. In: Machado, P., McDermott, J., Carballal, A. (eds.) Evolutionary and Biologically Inspired Music, Sound, Art and Design, pp. 133–144. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36955-1_12

5. Ghosh, R., et al.: Robust inverse framework using knowledge-guided self-supervised learning: an application to hydrology. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 465–474 (2022)

6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)

7. Hamilton, D.P., Schladow, S.G.: Prediction of water quality in lakes and reservoirs. part i-model description. Ecol. model. **96**(1-3), 91–110 (1997)

8. Hipsey, M.R., et al.: A general lake model (GLM 3.0) for linking with high-frequency sensor data from the global lake ecological observatory network (GLEON). Geoscientific Model Dev. **12**(1), 473–523 (2019)

9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

10. Janssen, A.B., Arhonditsis, G.B., Beusen, A., Bolding, K., Bruce, L., Bruggeman, J., Couture, R.M., Downing, A.S., Alex Elliott, J., Frassl, M.A., et al.: Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective. Aquat. Ecol. **49**, 513–548 (2015)

11. Jenny, J.P., et al.: Urban point sources of nutrients were the leading cause for the historical spread of hypoxia across European lakes. Proc. Nat. Acad. Sci. **113**(45), 12655–12660 (2016)

12. Jia, X., et al.: Physics guided RNNs for modeling dynamical systems: a case study in simulating lake temperature profiles. In: Proceedings of the 2019 SIAM International Conference on Data Mining, pp. 558–566. SIAM (2019)

13. Khawar, F., Hang, X., Tang, R., Liu, B., Li, Z., He, X.: AutoFeature: searching for feature interactions and their architectures for click-through rate prediction. In: ACM International Conference on Information and Knowledge Management (CIKM), pp. 625–634 (2020)

14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

15. Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. **23**(12), 5089–5110 (2019)

16. Ladwig, R., et al.: Long-term change in metabolism phenology in north temperate lakes. Limnol. Oceanogr. **67**(7), 1502–1521 (2022)

17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)

18. Li, J., et al.: Feature selection: a data perspective. ACM comput. surv. (CSUR) **50**(6), 1–45 (2017)

19. Li, X., Epitropakis, M.G., Deb, K., Engelbrecht, A.: Seeking multiple solutions: an updated survey on niching methods and their applications. IEEE Trans. Evol. Comput. **21**(4), 518–538 (2016)

20. Lin, J.Y., Ke, H.R., Chien, B.C., Yang, W.P.: Classifier design with feature selection and feature extraction using layered genetic programming. Expert Syst. Appl. **34**(2), 1384–1393 (2008)

21. Liu, B., Xue, N., Guo, H., Tang, R., Zafeiriou, S., He, X., Li, Z.: AutoGroup: Automatic feature grouping for modelling explicit high-order feature interactions in CTR prediction. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 199–208 (2020)

22. Liu, B., Zhu, C., Li, G., Zhang, W., et al.: AutoFIS: automatic feature interaction selection in factorization models for click-through rate prediction. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), pp. 2636–2645 (2020)

23. Liu, H., Simonyan, K., Yang, Y.: Darts: differentiable architecture search. In: International Conference on Learning Representations (2019)

24. Liu, Y., et al.: A survey on evolutionary neural architecture search. IEEE Trans. Neural Netw. Learn. Syst. **34**(2), 550–570 (2021)

25. Meyer, M.F., et al.: National-scale remotely sensed lake trophic state from 1984 through 2020. Sci. Data **11**(1), 77 (2024)

26. Muni, D.P., Pal, N.R., Das, J.: Genetic programming for simultaneous feature selection and classifier design. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **36**(1), 106–117 (2006)

27. Phillips, J.S.: Time-varying responses of lake metabolism to light and temperature. Limnol. Oceanogr. **65**(3), 652–666 (2020)

28. Saloranta, T.M., Andersen, T.: MyLake-a multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. Ecol. Model. **207**(1), 45–60 (2007)

29. Shang, R., et al.: A multi-population cooperative coevolutionary algorithm for multi-objective capacitated arc routing problem. Inf. Sci. **277**, 609–642 (2014)

30. Solomon, C.T., et al.: Ecosystem respiration: drivers of daily variability and background respiration in lakes around the globe. Limnol. Oceanogr. **58**(3), 849–866 (2013)

31. Sommer, U., et al.: Beyond the plankton ecology group (peg) model: mechanisms driving plankton succession. Annu. Rev. Ecol. Evol. Syst. **43**, 429–448 (2012)

32. Song, W., et al.: Autoint: automatic feature interaction learning via self-attentive neural networks. In: ACM International Conference on Information and Knowledge Management (CIKM), pp. 1161–1170 (2019)

33. Staehr, P.A., et al.: Lake metabolism and the diel oxygen technique: state of the science. Limnol. Oceanogr. Methods **8**(11), 628–644 (2010)

34. Tang, K., Yang, P., Yao, X.: Negatively correlated search. IEEE J. Sel. Areas Commun. **34**(3), 542–550 (2016)

35. Telikani, A., Tahmassebi, A., et al.: Evolutionary machine learning: a survey. ACM Comput. Surv. **54**(8), 1–35 (2021)

36. Weber, M., Neri, F., Tirronen, V.: Shuffle or update parallel differential evolution for large-scale optimization. Soft. Comput. **15**(11), 2089–2107 (2011)

37. Wilson, P.C.: Water quality notes: dissolved oxygen: Sl313/ss525, 1/2010. EDIS **2010**(2) (2010)

38. Woolway, R.I., et al.: Phenological shifts in lake stratification under climate change. Nat. Commun. **12**(1), 2318 (2021)

39. Wu, G., Mallipeddi, R., Suganthan, P.N., Wang, R., Chen, H.: Differential evolution with multi-population based ensemble of mutation strategies. Inf. Sci. **329**, 329–345 (2016)

40. Xiao, L.: Dual averaging method for regularized stochastic learning and online optimization. Adv. Neural Inf. Proc. Syst. **22** (2009)

41. Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. IEEE Trans. Evol. Comput. **20**(4), 606–626 (2015)
42. Ye, Y., et al.: Mane: organizational network embedding with multiplex attentive neural networks. IEEE Trans. Knowl. Data Eng. **35**(4), 4047–4061 (2022)
43. Yu, R., Xu, X., Ye, Y., Liu, Q., Chen, E.: Cognitive evolutionary search to select feature interactions for click-through rate prediction. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3151–3161 (2023)
44. Yu, R., Ye, Y., Liu, Q., Wang, Z., Yang, C., Hu, Y., Chen, E.: XCrossNet: feature structure-oriented learning for click-through rate prediction. In: Advances in Knowledge Discovery and Data Mining: 25th Pacific-Asia Conference (PAKDD), pp. 436–447. Springer (2021). https://doi.org/10.1007/978-3-030-75765-6_35
45. Zhang, F., Mei, Y., Zhang, M.: A two-stage genetic programming hyper-heuristic approach with feature selection for dynamic flexible job shop scheduling. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 347–355 (2019)
46. Zhao, H., Wu, X., et al.: CoEA: a cooperative-competitive evolutionary algorithm for bidirectional recommendations. IEEE Trans. Evol. Comput. **26**(1), 28–42 (2021)
47. Zhu, L., Ma, Y., Bai, Y.: A self-adaptive multi-population differential evolution algorithm. Nat. Comput. **19**, 211–235 (2020)