


Research Article

Application Research of Petroleum Basic Data Mining System Based on Intelligent Computing and Decision Tree Algorithm

Xiaoyan Wang ¹, Jinjin Liang,¹ Xiaoming He,² and Hua Yang³

¹School of Science, Xi'an Shiyou University, Xi'an, 710065 Shaanxi, China

²CNOOC International Limited, Iraq Overseas Branch, Beijing 100028, China

³CNPC Changqing Oilfield Company, Xi'an, 710021 Shaanxi, China

Correspondence should be addressed to Xiaoyan Wang; wangwang@xsyu.edu.cn

Received 19 April 2022; Revised 27 June 2022; Accepted 8 July 2022; Published 25 July 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Xiaoyan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent improvements in data mining technologies, besides the IoT, enable the implementation of a strategy for boosting oil output from oil wells. As a regularly employed improved oil recovery technology, steam flood injection takes use of thermodynamic and gravitational capabilities to deploy and neutralize oil on-site to raise oil output. Instead of relying on conventional physics to model steam floods, this research proposes using a combination of a chimp optimization algorithm (ChOA) and a decision tree to better represent steam flood performance. We present a method for dealing with a particular type of petroleum time series data using ChOA in conjunction with decision trees and IoT. It is shown that the method is useful in predicting oil production in steam floods. Even more impressive is the 4.02 percent increase in oil output that may be achieved via the use of a new optimization system that offers the best possible steam allocation plan. Our objective has been to develop a cloud-based minimum viable product capable of data collection and storage and also training and deployment of a cloud ChOA model. Predictive maintenance, for example, might benefit from this workflow's ability to analyze time series data.

1. Introduction

IoT and intelligent systems both have a solid track record in a variety of sectors. The petroleum industry has recently begun to pay greater attention to them. Traditional petroleum industry difficulties can be solved using an intelligent system. The applications of computer science in geoscience were outlined in reference [1]. Using computer vision, petrophysics has made significant progress over the last few years.

Real-time data collection and model generation can be performed through the establishment of IoT networks or a cloud hosting. The petroleum industry exploration and production use it extensively, notably in the upstream sector. It has a wide range of uses. This IoT-based architecture is described in detail in reference [2], which both provide an overview of recent advancements in sensor networks for the petroleum industry as well as open problems that have to be addressed. In order to forecast a possible three percent increase under oil production, an intelligent system and the Internet of Things were used to establish an effective method

for predicting oil production in steam flood situations and an optimization system for steam assignment.

Research into ways to increase oil output in order to fulfill rising global energy needs is a hot topic. One of the most often utilized techniques for increasing postnatural oil extraction oil output is enhanced oil recovery. Due to low well pressure, the natural pumping step often results in 70% residual crude oil. There are three basic ways of increased oil recovery: heat injection, gas injection, and chemical injection. When heavy oil is mobilized and diluted utilizing steam and gravity potential, production wells may readily recover oil from reservoirs by injecting steam into infill wells.

For decades, experts have been trying to predict how much oil steam flood fields will produce. Steam flood performance and oil output may be predicted using traditional analytical models, which were created utilizing physics concepts as well as reservoir parameters. Contrary to expectations, real oil output was far lower than predicted. Steam flood injection intelligent system research is few and far between. A novel steam flood screening criterion was

developed in reference [3], using the clustering technique, an intelligent system approach, in order to assist in the selection of the improved oil recovery method to be used for various reservoir circumstances. However, it does not include projections for oil output.

It has been claimed in recent years that heavy oil wells can be improved by using a decision tree model [4]. However, decision trees need knowledge of the properties of the optimization issue and some information about the gradient. In recent years, the use of metaheuristic algorithms (MAs) in optimization problems has grown in popularity [5, 6].

According to [7–9], metaheuristics can be divided into two main classes: single-solution-based and population-based. In the former class (simulated annealing for instance), the search process starts with one candidate solution. This single candidate solution is then improved over the course of iterations. Population-based metaheuristics, however, perform the optimization using a set of solutions (population). In this case, the search process starts with a random initial population (multiple solutions), and this population is enhanced over the course of iterations. Population-based metaheuristics have some advantages compared to single-solution-based algorithms:

- (i) Multiple candidate solutions share information about the search space which results in sudden jumps toward the promising part of search space
- (ii) Multiple candidate solutions assist each other to avoid locally optimal solutions
- (iii) Population-based metaheuristics generally have greater exploration compared to single-solution-based algorithms

One of the interesting branches of the population-based metaheuristics is swarm intelligence algorithm (SIA).

However, in this paper, we try to divide the metaheuristic algorithm by their nature-inspiring origin, as previous authoritative references have made these categories in other ways [10–13, 16]. In this kind of categorization, there are both single-vector and swarm methods in each category. For example, in the physic-based category, there is GA (population-based) and SA (single-solution-based).

NFL theorem [14] states that there is no MA that can effectively address all optimization issues as the best approach. This has resulted in the creation of novel MAs capable of addressing a range of optimization problems. The ChOA is a new tool for collective hunting that mimics the expertise and sexual encouragement of agents. According to [15], this algorithm has the potential to outperform other MAs in terms of performance. Researchers have used the ChOA algorithm in three different types of studies since it was first introduced in 2020.

In general, the following are some advantages that ChOA has over other MOAs:

- (i) Especially for situations involving larger dimensions, the division of chimps into autonomous

groups ensures that the search space will be thoroughly explored

- (ii) ChOA exploitability is highlighted by the hypothesized semideterministic feature in chaotic maps
- (iii) Chaotic maps aid the ChOA method in resolving local stagnations of optimas
- (iv) As a result of the ChOA algorithm's use of four different populations of search agents, local optima avoiding is extremely high
- (v) As the number of iterations rises, special forms of certain parameters encourage exploitation and convergence
- (vi) Chimps learn about the search space with each iteration
- (vii) In order to preserve the greatest answer so far, ChOA nearly relies on its own memory
- (viii) ChOA contains a few settings that can be adjusted
- (ix) The suggested approach is simple to implement due to the parallel nature of independent (unrelated) groups and the flexibility of ChOA
- (x) Chimpanzees have a wide range of abilities and knowledge, but they all work together as part of a group of hunters. As a result, each hunter's unique skill set can be put to good use at different points in the hunt

These include time series prediction [16], COVID-19-positive cases' detection by X-ray pictures [15], economic load dispatching, efficient fuzzy classification [17], and sonar database categorization [10]. Despite the fact that these studies have some validity, attempting to solve a well-known problem by introducing new paradigms or approaches is not a productive research strategy.

Other optimization methods, such as hybrid SCA-ChOA [18], combination random vector functional ChOA, and spotted hyena Sh-ChOA, are employed in conjunction with ChOA to increase their performance [19]. For long-term, low-power underwater sensor networks, a hybrid ChOA/HGS approach has been proposed [13]. Hybrid models may be more accurate, but their excessive complexity makes them inappropriate for difficult situations, especially those with several dimensions.

Finally, academics have attempted to increase the ChOA's performance by creating or altering specific operators. In order to speed up ChOA convergence, the WChOA employed an average weighting approach [8]. Niching-ChOA adopted ChOA's niching approach in order to boost its exploratory abilities [11], whereas the EChOA [9] used the incredibly destructive exponential mutation and correlation parameters to start the population with the lowest social status chimpanzees. Classifiers were developed using Fuzzy-ChOA [20], which uses fuzzy models to modify the ChOA. DLF-ChOA (dynamic levy flight ChOA) is a variation on ChOA that aims to improve global performance [7].

Given the novelty of ChOA, some study has focused on how it may be improved. Because of this, its precision and convergence rate may be enhanced even more. The goal of this research is to increase the decision tree's performance by using it in conjunction with ChOA for predicting oil production in steam floods.

According to the following structure, the rest of this paper will be as follows: Section 2 introduces the subject matter, Section 3 explains the technique, Section 4 gives the simulation results and discussion, and Section 5 concludes the article.

2. Background Materials

The decision tree and ChOA will be briefly reviewed in this section.

2.1. Decision Tree. A decision tree is just a data flow diagram architecture where each node in the middle represents tests and each route represents the results of the tests. This means that each leaf node represents a class label. The roots and leaves of a tree indicate classification criteria. Visual and analytical help for decision-making is provided by decision trees and influence diagrams, in which the predicted values (or usefulness) of competing choices are estimated. Three types of nodes may be found in a decision tree's branches:

- (i) Decision nodes: a typical way to represent these nodes is to use squares
- (ii) Nodes of randomness: usually depicted by concentric circles
- (iii) End nodes: end nodes are frequently denoted with triangles

Operation research frequently employs decision trees. A probability model should be used as the best choice framework or online selection model method if decision-making should be carried out online without any recall under incomplete knowledge. Decision trees may also be used to calculate conditional probabilities by providing a visual representation of the relationships between variables.

2.2. Chimp Optimization Algorithm. A fission-fusion system governs chimp social life. The mix of society in this type of civilization changes with time. For each community member, there is an inherent talent or task that might vary throughout time [21]. Due to the fact that each chimp group has its own unique capacity to do a certain task, the notion of independent groups is included in this algorithm.

According to prior research, chimps may be divided into four categories: drivers, barriers, chasers, and attackers. To ensure a successful hunt, they are tasked with a variety of responsibilities. Rather than attempting to seize their victim, drivers simply follow along behind them. The prey's escape route is obstructed by barriers erected in trees. Chasers pursue their prey at breakneck speed in order to snatch them. As a final precaution, predators know the prey's escape path down into the lower treetops. Figure 1 depicts the various stages of the hunt. Attackers must be better at guessing the prey's future

moves. As a result, a successful hunt is rewarded with a bigger portion of meat for the attacker. Age, intelligence, and physical prowess all play a part in a player's ability to attack. It is possible for chimpanzees to switch roles during a hunt or retain their roles throughout the entire operation.

According to a recent study, chimpanzees seek meat in order to exchange it for social privileges like sex or grooming. Because of this, intelligence and knowledge may have an unintentional influence on the quest. The term "social incentives" has been used solely by humans and chimpanzees, to our knowledge. A unique advantage is given to chimpanzees in this regard compared to other social predators. As a result, chimpanzees operate in a disorderly fashion in the closing stages of the hunt, allowing each chimpanzee to pursue his or her own goal of obtaining prey. Chimps hunt in two phases: "exploration," in which they drive, block, and chase their prey, and "exploitation," in which they assault it. Exploitation is the first step of chimpanzee social hunting. Figure 1 depicts these two stages. In the next part, we will go through the mathematical models for the hunt's first two stages and its last four steps [21].

As previously stated, hunting occurs during the exploratory and exploitative stages. Equations (1) and (2) are used to numerically simulate driving and pursuing the prey [21].

$$\mathbf{d} = |\mathbf{c} \cdot \mathbf{x}_{\text{prey}}(t) - \mathbf{m} \cdot \mathbf{x}_{\text{chimp}}(t)|, \quad (1)$$

$$\mathbf{x}_{\text{chimp}}(t+1) = \mathbf{x}_{\text{prey}}(t) - \mathbf{a} \cdot \mathbf{d}, \quad (2)$$

where t is the latest iteration, \mathbf{a} , \mathbf{m} , and \mathbf{c} are the regression coefficients, and $\mathbf{x}_{\text{chimp}}$ and \mathbf{x}_{prey} denote the chimp and prey position vectors, respectively. Additionally, \mathbf{a} , \mathbf{m} , and \mathbf{c} coefficients are determined using

$$\mathbf{a} = 2 \cdot \mathbf{f} \cdot \mathbf{r}_1 - \mathbf{a}, \quad (3)$$

$$\mathbf{c} = 2 \cdot \mathbf{r}_2, \quad (4)$$

$$\mathbf{m} = \text{Chaotic_value}. \quad (5)$$

The iteration procedure reduces \mathbf{f} nonlinearly from 2.5 to 0. Additionally, \mathbf{r}_1 and \mathbf{r}_2 are random vectors inside the range [0,1]. Furthermore, \mathbf{m} is a chaotic variable computed from the numerous chaotic maps; this vector therefore captures the chimps' desire on the hunting process. The next sections will provide a detailed description of these vectors. The classic population-based method allows individuals to be seen as a single group with a common search strategy since all individuals act identically in local and global searches. It is possible to obtain a straight and a stochastic search result using different population-based techniques, but, this is only theoretical. Independent chimp groups will be mathematically described in the next sections, in which several methods of updating \mathbf{f} are used. To keep the separate groups up to date, you can use any continuous function. \mathbf{f} must be reduced with each iteration of these equations. It must be noted that the coefficient vectors are indicated in bold type such as \mathbf{m} , \mathbf{a} , and \mathbf{d} , and every multiplication between two arrays such as " $\mathbf{a} \cdot \mathbf{d}$ and $\mathbf{f} \cdot \mathbf{r}_1$ " means array multiplication.

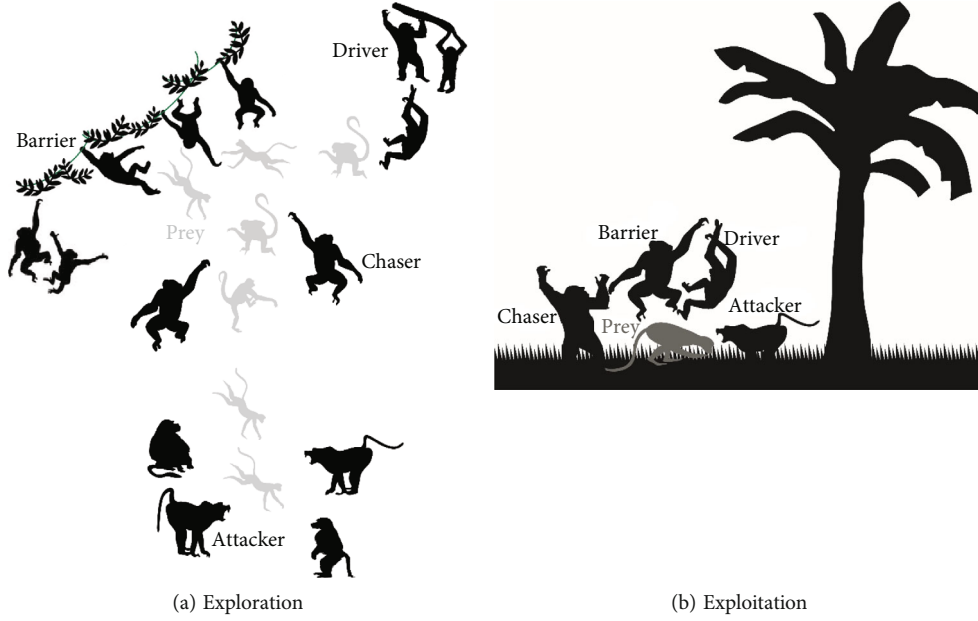
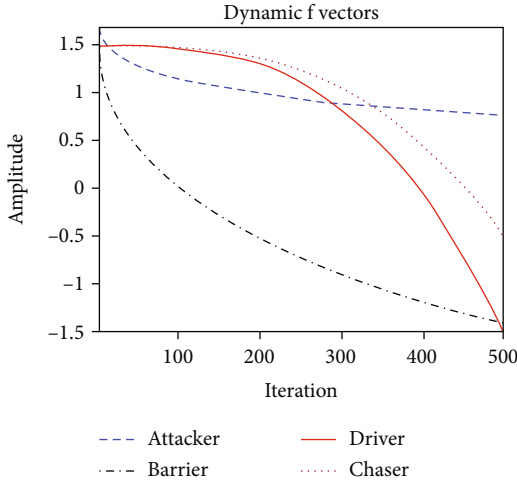


FIGURE 1: Two main stages of ChOA.

TABLE 1: The f vectors.

Groups	Driver	Attacker	Chaser	Barrier
f	$(-3t^3/T^3) + 1.5$	$1.95 - 2t^{1/4}/T^{1/3}$	$(-2t^3/T^3) + 1.5$	$1.95 - 2t^{1/3}/T^{1/4}$

FIGURE 2: Mathematical models of the nonlinear reduction of f vectors.

All minimization strategies face the challenge of locating the global minimum. In population-based optimization approaches, generally, the preferred manner to converge towards the global solution can be divided into two main stages (exploration and exploitation).

It is important to encourage individuals to spread out during the initial phases of optimization. In other words, instead of focusing on local minima, they should strive to examine the entire search space. In the later steps, the indi-

viduals have to exploit knowledge obtained to converge on the global solution. In ChOA, by fine-adjusting of the variable f , we may combine these two stages to discover global minimum with rapid convergence speed. Based on these considerations, we propose the concept of autonomous groups called “attacker, barrier, chaser, and driver,” which is a four-part system.

Individual particles search the issue space using their own technique, which is tuned by f . This means that particles can be viewed as a single group with a single strategy when it comes to both local and global search in population-based optimization methods. However, a population-based optimization algorithm might possibly result in a more randomized and directed search at the same time if it uses a variety of autonomous groups with a shared goal. The autonomous groups are mathematically modeled in this study, using several methods for updating f . To put it another way, the organizations differ in their approach to exploration and exploitation. An autonomous group’s strategy can be updated using any continuous function that has a range between $[0, L]$.

These four autonomous groups each employ their own pattern to investigate the search space on a local and global scale. Additionally, the most effective variations of ChOA with distinct subgroups are selected from among the numerous techniques evaluated. The dynamic features of f are depicted in Table 1 and Figure 2. t signifies the current iteration, while T denotes the total number of iterations possible. In order to increase the quality of ChOAs, dynamic

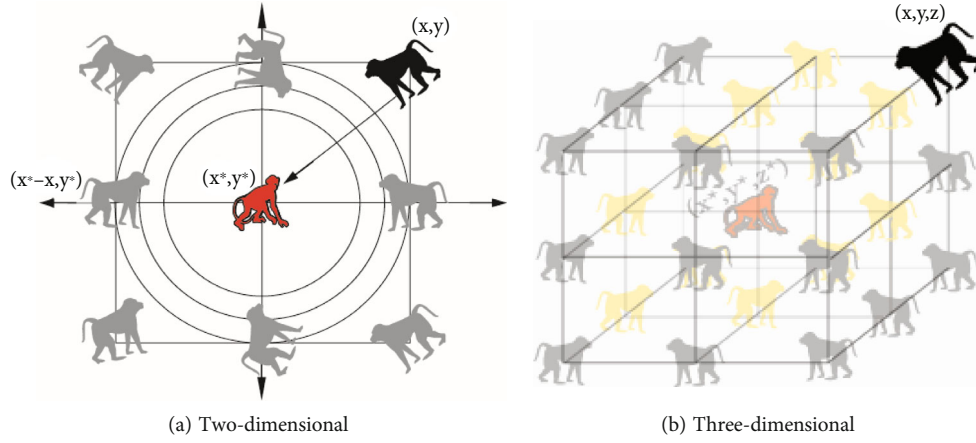


FIGURE 3: The 2D and 3D presentation of chimps' positions.

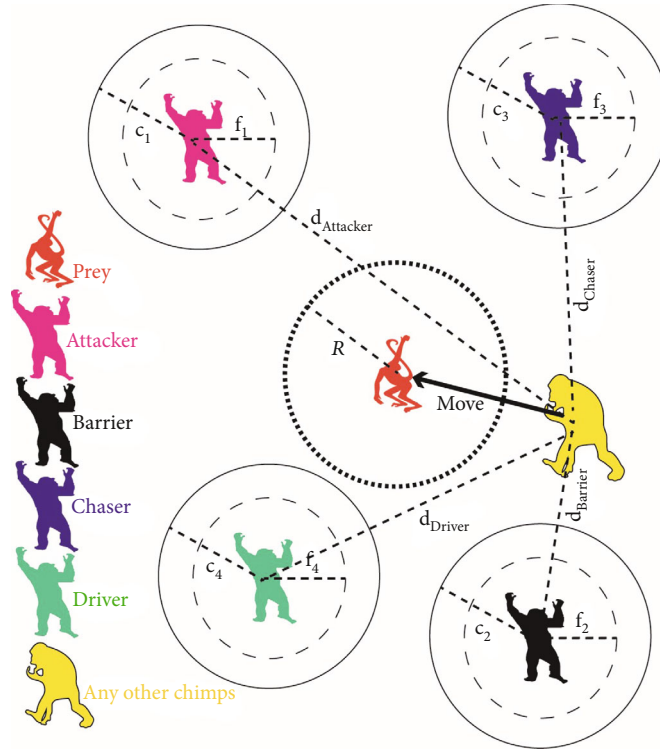


FIGURE 4: The producers of chimp's updating position.

equations were adjusted to include a wide variety of curves and slopes, guaranteeing that each group shows a different exploration behavior [21].

To illustrate the notion of Equations (1) and (2), Figure 3 depicts a 2D and 3D representation of a chimp, as well as a number of its probable future places. As shown, a chimpanzee in location (x, y) can adjust its location in relation to the location of its prey (x^*, y^*) by modifying the values of the \mathbf{a} and \mathbf{c} coefficients. It is worth noting that the chimpanzees have access to every point in the search area via the randomized variables \mathbf{r}_1 and \mathbf{r}_2 . This method may be generalized to search an n -dimensional space.

To arithmetically design chimps attacking behavior patterns, two approaches are taken: in the first place, the chim-

panzees are able to find their prey's position by driving, obstructing, and chasing it. Finally, attackers are often in charge of the hunt. The optimal posture of the prey during the initial iteration is unknown. To address this problem, the attacker's location is considered to be that of the prey. As a result, the optimal solution are preserved, and other chimpanzees are prompted to adjust their locations depending on the best chimps' placements. Equations (6) to (8) define this approach [21]:

$$\begin{aligned} \mathbf{d}_A &= |\mathbf{c}_1 \cdot \mathbf{x}_A - \mathbf{m}_1 \cdot \mathbf{x}|, \mathbf{d}_B = |\mathbf{c}_2 \cdot \mathbf{x}_B - \mathbf{m}_2 \cdot \mathbf{x}|, \mathbf{d}_{Ch} \\ &= |\mathbf{c}_3 \cdot \mathbf{x}_{Ch} - \mathbf{m}_3 \cdot \mathbf{x}|, \mathbf{d}_D = |\mathbf{c}_4 \cdot \mathbf{x}_D - \mathbf{m}_4 \cdot \mathbf{x}|, \end{aligned} \quad (6)$$

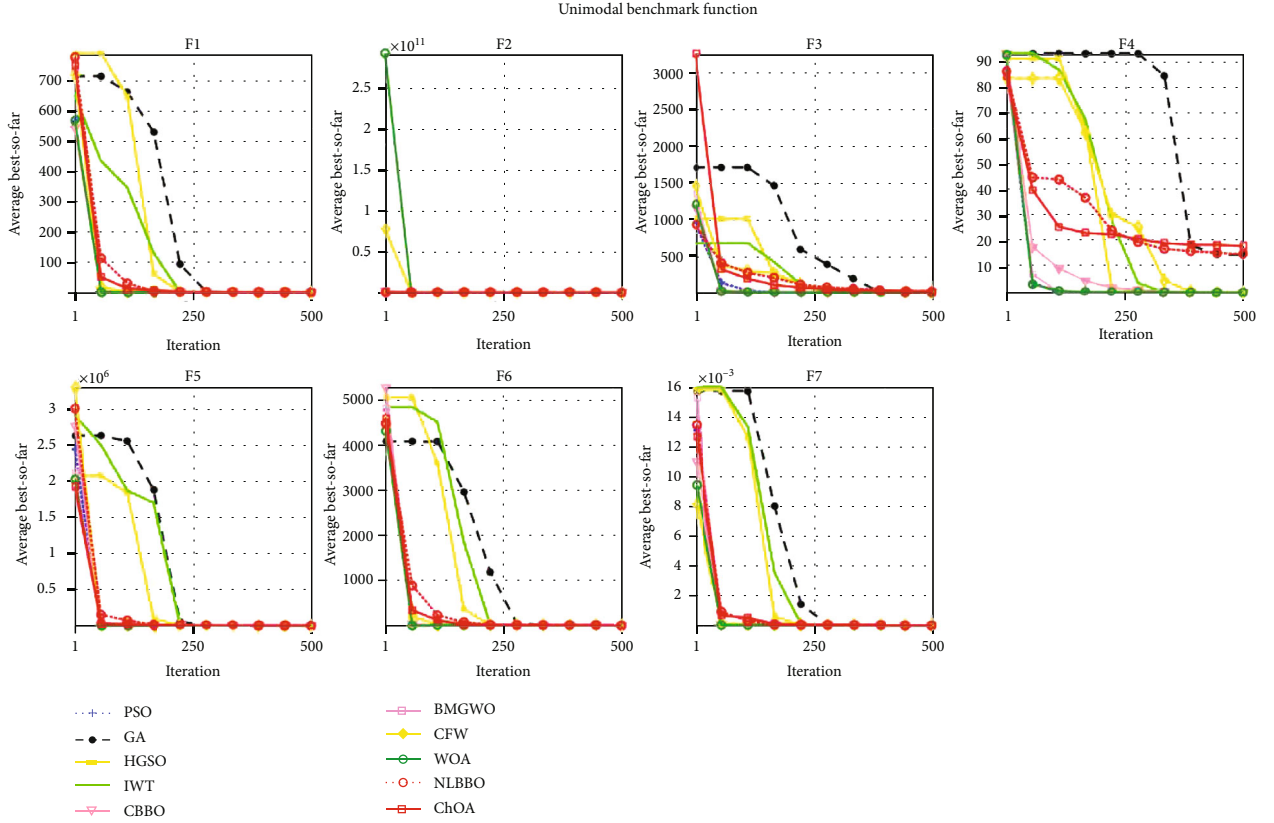
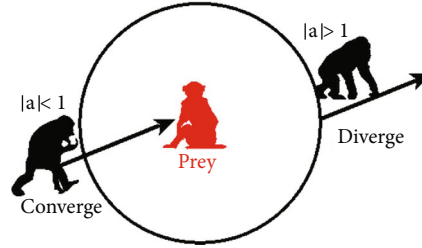


FIGURE 5: The algorithms' convergence curves in unimodal function.

FIGURE 6: A chimp's positional update system is affected by \mathbf{a} .

$$\begin{aligned}
 \mathbf{x}_1 &= \mathbf{x}_A - \mathbf{a}_1(\mathbf{d}_A), \mathbf{x}_2 = \mathbf{x}_B - \mathbf{a}_2(\mathbf{d}_B), \mathbf{x}_3 = \mathbf{x}_{Ch} - \mathbf{a}_3(\mathbf{d}_{Ch}), \mathbf{x}_4 \\
 &= \mathbf{x}_D - \mathbf{a}_4(\mathbf{d}_D),
 \end{aligned} \quad (7)$$

$$\mathbf{x}(t+1) = \frac{1}{4} \sum_{k=1}^4 \mathbf{x}_k. \quad (8)$$

A chimpanzee's location in the search area is constantly changed based on the position of other chimpanzees, as seen in Figure 4. The chimpanzee's ultimate location may be observed to be in a circle established by the positions of the assailant, barrier, pursue, and drivers.

To summarize, the chimps will assault their prey until it is no longer moving and then end their search. Reducing the value of \mathbf{f} linearly will help us better understand how an assault works. The range of \mathbf{a} vector is likewise lowered, much like the range of \mathbf{f} vector. As seen in Figure 5, a

chimpanzee's future position will be anywhere between its current position and the position of the prey. Despite the recommended pushing, blocking, and pursuing tactics, there is still a possibility that chimps will be trapped in local minima. To emphasize the exploratory skills in the evaluation stage, another operator is required. Instead of relying on only one operator, ChOA requires an additional operator during the exploitation phase to help avoid local minima trapping.

Divergent and converging chimpanzee attacks are common in ChOA. This behavior is mathematically represented in Figure 6 by allocating vector \mathbf{a} .

ChOA's exploration phase gets a boost with the addition of the \mathbf{c} vector. Equation (4) demonstrates that \mathbf{c} is a random vector with a range of $[0, 2]$. Predator-prey distance may be calculated mathematically using Equation (4). When chimps are out in the wild, the vector of \mathbf{c} acts as a deterrent, keeping them from getting close to their prey. Prey can be made more difficult or easy to catch thanks to the vector of \mathbf{c} .

TABLE 2: Chaotic maps.

No.	Name	Chaotic map	Range
1	Quadratic	$x_{i+1} = x_i^2 - c$	(0,1)
2	Chebyshev	$x_{i+1} = \cos(i \times \cos^{-1}(x_i))$	(-1,1)
3	Gauss/mouse	$x_{i+1} = \begin{cases} 1, & x_i = 0 \\ 1/\text{mod}(x_i, 1), & \text{else} \end{cases}$	(0,1)
4	Iterative	$x_{i+1} = \sin(7\pi/10 \times x_i)$	(-1,1)
5	Logistic	$x_{i+1} = \alpha x_i(1 - x_i), \alpha = 4$	(0,1)
6	Singer	$x_{i+1} = (7.75x_i - 23.29x_i^2 + 28.64x_i^3 - 13.29x_i^4) \times \mu$	(0,1)
7	Bernoulli	$x_{i+1} = 2 \times x_i \pmod{1}$	(0,1)
8	Sine	$x_{i+1} = \sin(\pi \times x_i)$	(0,1)
9	Tent	$x_{i+1} = \begin{cases} 10 \times x_i/7, & x_i < 0.7 \\ 10/3(1 - x_i), & 0.7 \leq x_i \end{cases}$	(0,1)
10	Circle	$x_{i+1} = \text{mod}(x_i + 0.2 - (1/4 \times \pi) \sin(2\pi x_k), 1)$	(0,1)

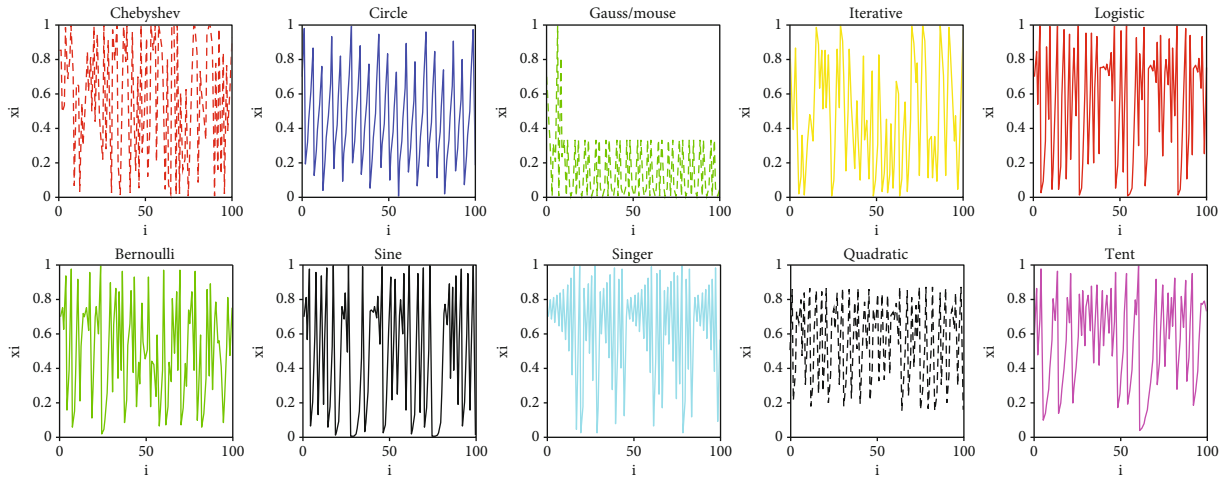


FIGURE 7: The chaotic map visualization.

Chimp's social motivation depends on hunting meat, as previously indicated. The chimpanzees are forced to give up their own hunts in order to get their hands on the final piece of hunting meat. Because of this, they will do everything to get their hands on flesh from a hunt in order to satisfy their social needs. Chaos maps can be used to simulate the chimpanzees' last stage behavior, which is characterized by randomness.

Maps utilized to increase ChOA's performance are shown in Table 2 and Figure 7. In reality, randomness may be generated by these deterministic processes as well. All chaotic maps have a starting value of 0.7. Modeling the update process in this way looks like this [21]:

$$\mathbf{x}_{\text{chimp}}(t+1) = \begin{cases} \mathbf{x}_{\text{prey}}(t) - \mathbf{a} \cdot \mathbf{d} & \text{if } \mu < 0.5, \\ \mathbf{Chaotic_value} & \text{if } \mu \geq 0.5, \end{cases} \quad (9)$$

where μ denotes a random number between 0 and 1. In this equation, the normal behavior of chimp for changing position $\mathbf{x}_{\text{chimp}}(t+1) = \mathbf{x}_{\text{prey}}(t) - \mathbf{a} \cdot \mathbf{d}$ is substituted by values from the chaotic map (**Chaotic_value**) to provide chaotic behaviors for justification of sexual motivation of chimp. Indeed, this term reduces the risk of getting stuck in local minima by changing the search space, chaotically. In fact, by using the chaotic maps, we can control how the search space is changed in addition to the random behavior.

Starting with a random chimpanzee population, ChOA is launched in the first phase of its development (candidate solutions). Four groups of chimps are randomly assigned. After that, each chimp uses its own group approach to update its \mathbf{f} coefficients. The prey's distance from each possible solution is then updated. Additionally, adaptive adjustment of the \mathbf{c} and \mathbf{m} results in the avoidance of local optima while increasing the pace of convergence. \mathbf{f} is also

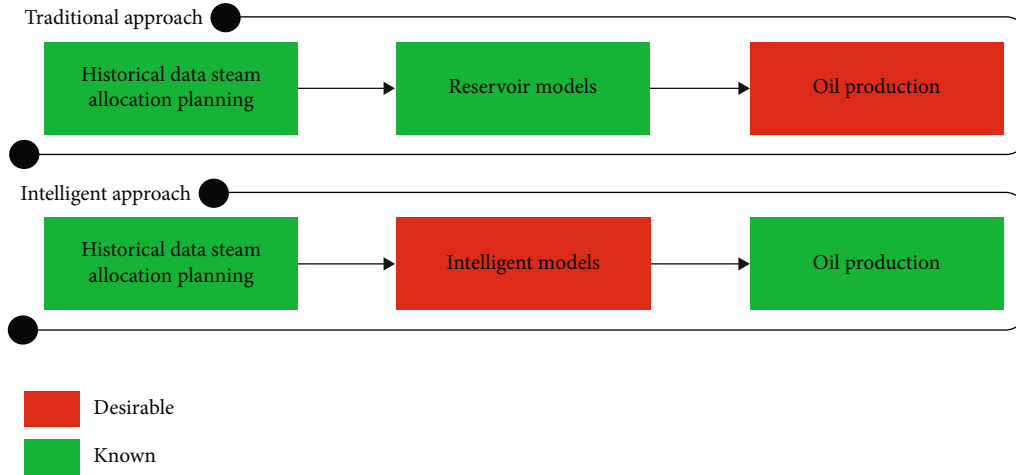


FIGURE 8: Comparison of intelligent and traditional approaches for oil production prediction.

TABLE 3: Structure of a well level's daily dataset.

Date	Identifier of well	Condition of well	Sensor data	Steam volume	Oil volume
2021-11-12	Prod	Pump	88	NA	19
2021-11-13	Prod	Shut-in	198	NA	29
2021-11-12	Infill	N/A	NA	5	N/A
2021-11-13	Infill	NA	NA	10	NA

TABLE 4: The initial values and parameters.

Value	Parameter	Algorithm
	f	Figure 1
ChOA	r_1, r_2	Random
	m	Table 1
	1	Immigration probability
NLBBO and CBBO	1	Max (I) and (E)
	0.005	The probability of mutation
HGSO	$\bar{v}c$	(0, 1]
BMGWO	a	Linearly decreased from 2 to 0
CFW	Velocity	5 m/s
	Initial velocity	3 m/s
	Total weight	10^{-3} kg
IWT	a	Linearly decreased from 2 to 0
10%	Elitism	
0.08	Mutation	GA
0.92	Crossover	
0.11	The inertial movement rate (α)	
0.65	C_1	PSO
0.93	C_2	

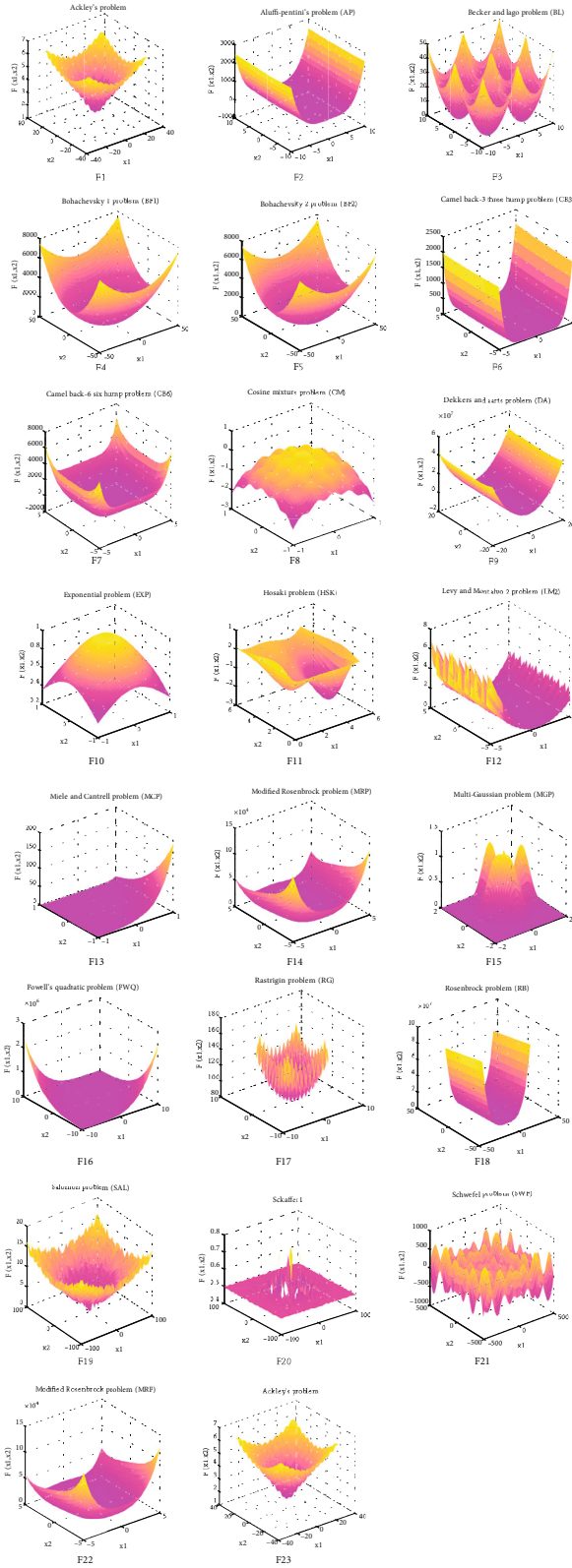


FIGURE 9: 3D plot of benchmark functions.

lowered to 0 from its previous value of 2.5, further aiding exploitation. Divergence from prey occurs when there is a disparity between predators and prey. Finally, chaotic maps aid in rapid convergence while avoiding local minima.

3. Model Developing

The classic technique relies on well-established algorithms and input data to forecast an output, such as oil production. This problem can be solved in a different way using intelligent system techniques. In the MA technique, candidate models are trained using the input and output from a training dataset (Figure 8).

The best model is determined using predetermined criteria and tested on a new dataset that was never used before. The conventional technique is motivated by physical principles, whereas data drive intelligent systems. Developing intelligent system models does not necessitate the use of geological parameters, as is the case with more traditional methods. The proportional value of each element in an intricate nonlinear system may be determined using decision tree-based intelligent system techniques such as random forests and XGBoost [22]. Using an intelligent system's tremendous capabilities, a steam flood monitoring team can statistically investigate ways to increase oil output. IoT and intelligent systems work well together. Because of the real-time data gathering, storage capabilities of IoT, and the ability to train and deploy intelligent system models on a cloud infrastructure or IoT devices, intelligent systems can analyze vast amounts of data gathered by IoT nodes to uncover previously unknown patterns. Incorporating IoT and intelligent system technologies is a win-win situation for this project.

3.1. Data Collection and Introduction. In the field, edge sensors gather raw data from five different data sources. Schematic design, principal keys, and sample frequency differ among the sources. To gather and cross-reference data, collect errors, and consolidate them into a single location, daily extract, transform, and load tasks are employed. After that, they have moved to a cloud storage facility so that the data engineering procedure may get started.

At the well level, Table 3 depicts the daily structure of a neighboring dependent well dataset. To accurately predict the daily oil output from each well, we want to create a single model for each pad. Sixteen variables are included in the dataset. Two types of wells fall under the well name umbrella: infill and production. Sensor data is a collection of real-time temperature and pressure readings taken in the field. There is a difference between infill wells and those that are currently producing oil: infill wells use steam volume as the daily amount of steam that is injected into the wells. An oil-producing well's daily output is known as its "oil volume." The empirical dataset is bound to have some missing data. The data science stage is required before putting data into a decision tree algorithm.

3.2. Methodology. There are five stages in the workflow process: data collection and transmission, statistical modeling, model creation, visualization of data, and system optimization. We use to the front and reverse copies to fill in the blanks for missing data for a variety of variables. Data engineering requires the whole history of all wells, so no records are discarded. The dataset has been divided into two subsets because there are two types of wells. For each of the infill

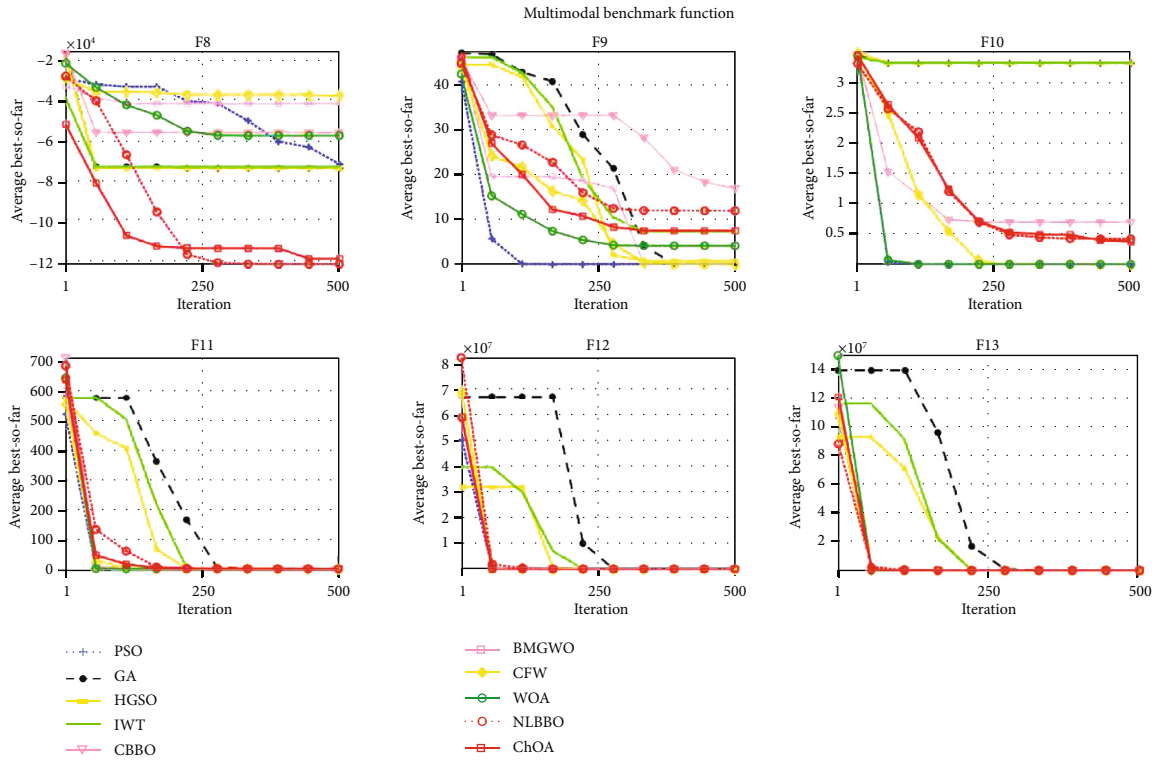


FIGURE 10: The algorithms' convergence curves in multimodal function.

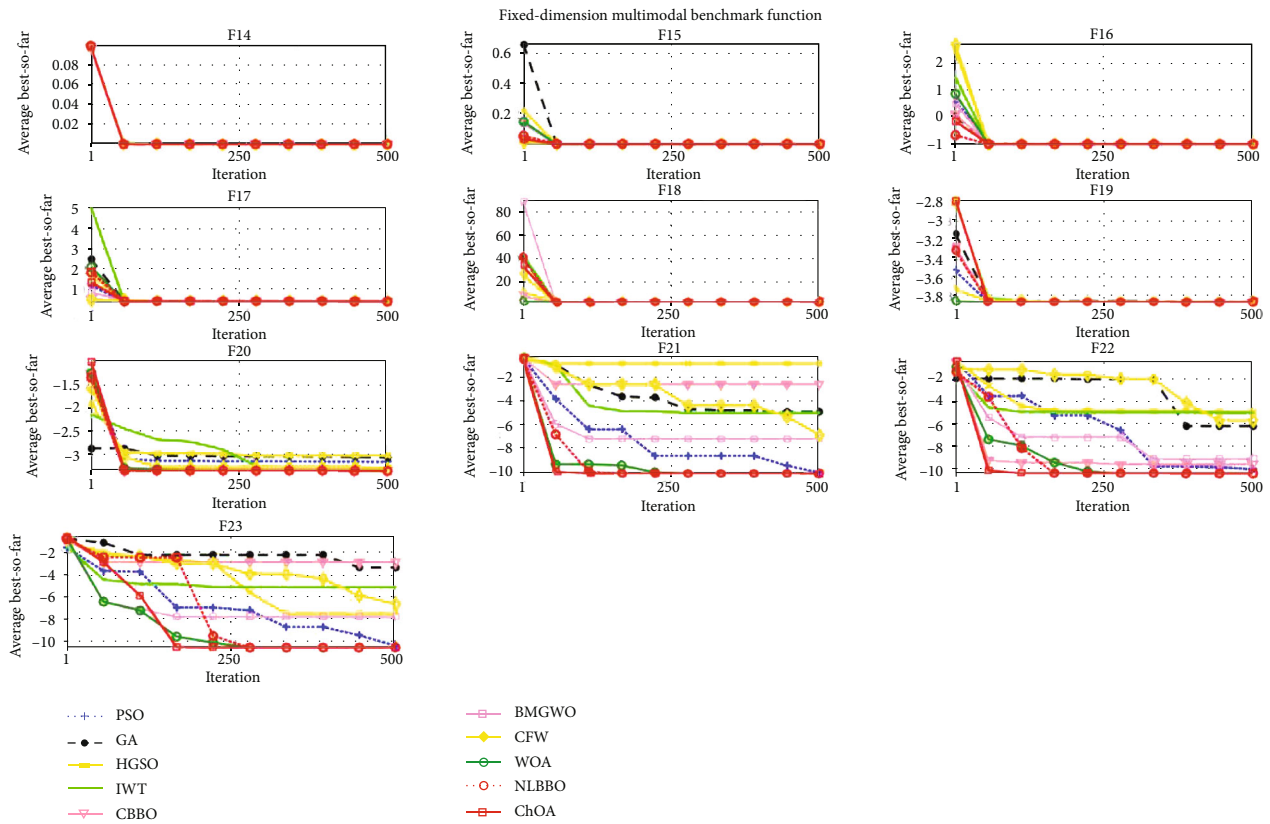


FIGURE 11: The algorithms' convergence curves in fixed-dimension function.

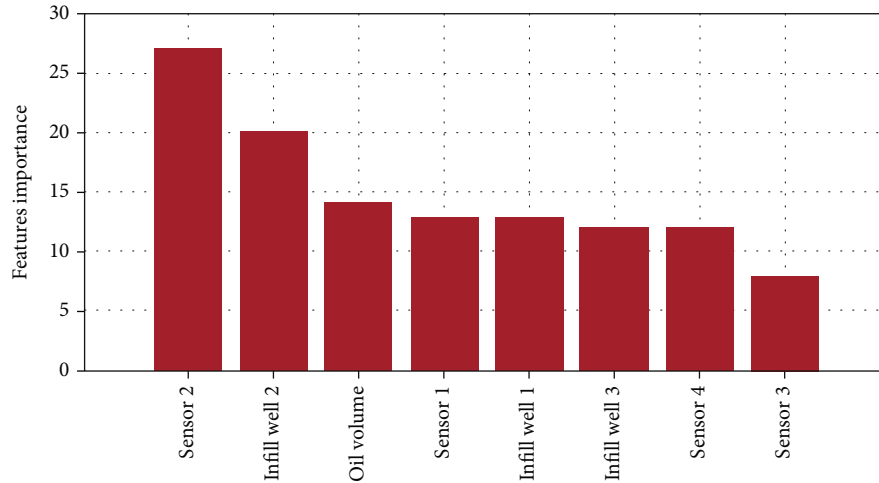


FIGURE 12: The model's eight most critical characteristics.

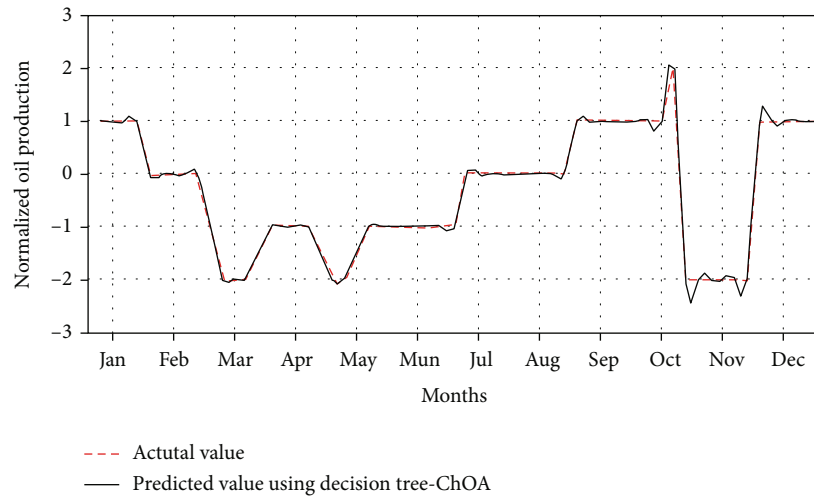


FIGURE 13: Comparison of actual values and predicted values using decision tree-ChOA for 12 months.

TABLE 5: The comparison of various benchmark algorithm in oil production prediction.

Algorithm	RMSE	R^2
ChOA	0.0012 ± 0.0023	0.852 ± 0.0134
CBBO	0.0213 ± 0.0055	0.633 ± 0.0425
NLBBO	0.0123 ± 0.0066	0.544 ± 0.0222
IWT	0.0152 ± 0.0045	0.725 ± 0.0352
PSO	0.0523 ± 0.0098	0.542 ± 0.0769
GA	0.0512 ± 0.0099	0.652 ± 0.0785
BMGWO	0.0265 ± 0.0045	0.733 ± 0.0211
HGSO	0.0285 ± 0.0044	0.722 ± 0.0178
CFW	0.0123 ± 0.0029	0.802 ± 0.0149

wells, we construct a new data format based on the daily amounts of steam injected into each row, column, and feature. Two sets of characteristics have been added to the

production-well subset. There is a gas/day/rate and a one-hot encoding of production wells based on categorical data, such as the well identifier and well condition. After reorganizing infill wells and aligning production wells by date, we create a new dataset by merging the two sets.

4. Experimental Results

On the benchmark functions, the suggested ChOA's performance is assessed first. After that, we will utilize the new ChOA model to get a better handle on oil output. Nine well-known algorithms, including PSO [23], GA, HGSO [24], IWT [25], CBBO [26], BMGWO [27], CFW [28], WOA [29], and NLBBO [30], were utilized to test ChOA's performance in benchmark functions. In MATLAB, each benchmark issue was run 30 times for each method. Table 4 also shows the algorithm's initial settings and parameters.

In this part, the suggested method is evaluated against 23 competitive benchmark functions. In order to determine the convergence trend and local optimums, we can use 16

multimodal functions and seven unimodal functions. Nine multimodal functions, as well as seven fixed dimension multimodal functions, are also utilized. In you will find further information. Benchmark difficulties are shown graphically in Figure 9. The comparison of various algorithms' convergence curves is shown in Figures 5–11.

Here, we provide the findings of a 30-day prediction model that we believe is the best. Figure 12 shows the model's eight most critical characteristics. In order to aid their planning and decision-making, a monitoring team can acquire a qualitative view of the importance of each aspect [29].

The dataset is separated into test (20 percent) and training (80 percent) datasets. According to Figure 13, real oil output is compared to a model's prediction of monthly oil production. To make things easier to see, the daily output and forecast are both calculated at the well levels and summed over all producing wells in a pad. The root mean square error (RMSE) measure is used to identify the best model from the fivefold cross-validation (RMSE). There has been a substantial improvement over prior studies in predicting test dataset results that are within the -10% to +10% range of the real outputs. As shown in Table 5, optimum and baseline models perform well on training and test datasets, respectively [30]. To anticipate future daily oil output, the baseline model uses the most recent 30-day previous daily production as a starting point. In terms of R^2 and RMSE, the decision tree-ChOA model exceeds the baseline model by a wide margin.

Optimizing steam flood distribution is simple when using a model that can reliably forecast oil production in various sensors. It is an example of a pad with triple infill wells, and infill well 2 took in 4211 m³ of steam volume every month, which is the true total monthly oil output. The model forecasts that 4242 m³ of oil will be produced, which is 0.9 percent more than the actual output. A brute-force search of all potential situations is the simplest way to find the best steam allocation strategy to optimize production. Because of the three infill wells, a set total steam volume is needed to calculate the steam volumes delivered into infill wells 1 and 2, which are two independent parameters.

For the best scenario, 30 percent steam injection into infill well 1, four percent injection into infill well 2, and 70 percent injection into infill well 3 yield a maximum oil output of 4339 m³, which is 3.3% higher than the actual production.

While brute-force search takes more and longer the more infill wells there are, it may be more practical to employ alternative optimization techniques, such as gradient descent search, in the future. In addition, this optimization system was created with the goal of increasing oil output. An alternative approach to optimization can be used if other goal functions, such as reducing the steam-to-oil ratio to reduce fuel costs, are defined.

In order to make this huge improvement, the search agent has been divided into four separate subgroups (assailant, obstacle, follower, and driver), and the chaos-based maps have been used rather than random numbers. As a result, the new method has no additional cost in terms of

processing power. This will only work if the various autonomous teams have been set up in concert. Otherwise, the ChOA's speed will be slightly reduced.

5. Conclusion

To handle a specific type of time series data, hybrid decision tree ChOA algorithm has been implemented in this research. In comparison to existing approaches, our model is able to estimate oil output in certain steam flood situations with an exceptional level of precision. The optimization system that we have built can also increase oil output by 3.21% by recommencing the best steam allocation plan. The emergence of cloud platforms for IoT has enabled us to create a cloud-based minimal level viable solution for steam flood optimization, allowing for real-time data collection, transmission, storage, and intelligent system training and deployment on cloud platforms. Steam floods may be studied more thoroughly thanks to this research. It would be fascinating to see how this technique may be applied to additional datasets with comparable time series structures. This method, for example, may be used to develop other metaheuristic algorithms to forecast the quantitative amount of abrasion on a machine part based on past records and the state of future operations.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflicts of Interest

The authors state that this article has no conflict of interest.

Acknowledgments

This work was supported by the General Project of Shaanxi Provincial Department of Science and Technology (2022JM-409), Research and Application of Intelligent Reservoir Dynamic Analysis Method.

References

- [1] K. N. Bhanu, H. J. Jasmine, and H. S. Mahadevaswamy, "Machine learning implementation in IoT based intelligent system for agriculture," *International Conference for Emerging Technology (INCET)*, vol. 2020, 5 pages, 2020.
- [2] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, M. S. Hossain, and M. Atiqzaman, "A reliable Internet of Things based architecture for oil and gas industry," in *In 2017 19th International conference on advanced communication Technology (ICACT)*, pp. 705–710, IEEE, PyeongChang, Korea (South), 2017.
- [3] M. Q. Hama, M. Wei, L. D. Saleh, and B. Bai, "Updated creening criteria for steam Flooding based on oil field projects data," in *In SPE Heavy Oil Conference-Canada*, Calgary, Alberta, Canada, 2014OnePetro.
- [4] M. Yan, J. C. MacDonald, C. T. Reaume, W. Cobb, T. Toth, and S. S. Karthigan, "Machine learning and the Internet of Things enable steam flood optimization for improved oil production," 2019, <http://arxiv.1908.11319>.

- [5] M. Azhdari, A. Mahmoodzadeh, and M. Khishe, "Power consumption optimization in underwater wireless sensor networks based on EECRU clustering algorithm for routing," *Iranian Journal of Marine Science and Technology*, vol. 24, no. 95, pp. 1–10, 2020.
- [6] J. Wu, M. Khishe, M. Mohammadi, S. H. T. Karim, and M. Shams, "Acoustic detection and recognition of dolphins using swarm intelligence neural networks," *Applied Ocean research*, vol. 115, article 102837, 2021.
- [7] W. Kaidi, M. Khishe, and M. Mohammadi, "Dynamic levy flight chimp optimization," *Knowledge-Based Systems*, no. - article 107625, 2021.
- [8] M. Khishe, M. Nezhadshahbodaghi, M. R. Mosavi, and D. Martín, "A weighted chimp optimization algorithm," *IEEE Access*, vol. 9, pp. 158508–158539, 2021.
- [9] J. Wang, M. Khishe, M. Kaveh, and H. Mohammadi, "Binary chimp optimization algorithm (BChOA): a new binary meta-heuristic for solving optimization problems," *Cognitive Computation*, vol. 13, no. 5, pp. 1297–1316, 2021.
- [10] M. Khishe and M. R. Mosavi, "Classification of underwater acoustical dataset using neural network trained by chimp optimization algorithm," *Applied Acoustics*, vol. 157, article 107005, 2020.
- [11] S.-P. Gong, M. Khishe, and M. Mohammadi, "Niching chimp optimization for constraint multimodal engineering optimization problems," *Expert Systems with Applications*, vol. 198, article 116887, 2022.
- [12] F. Chen, C. Yang, and M. Khishe, "Diagnose Parkinson's disease and cleft lip and palate using deep convolutional neural networks evolved by IP-based chimp optimization algorithm," *Biomedical Signal Processing and Control*, vol. 77, article 103688, 2022.
- [13] Y. Yang, Y. Wu, H. Yuan, M. Khishe, and M. Mohammadi, "nodes clustering and multi-hop routing protocol optimization using hybrid chimp optimization and hunger games search algorithms for sustainable energy efficient underwater wireless sensor networks," *Sustainable Computing: Informatics and Systems*, vol. 35, article 100731, 2022.
- [14] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [15] T. Hu, M. Khishe, M. Mohammadi, G.-R. Parvizi, S. H. T. Karim, and T. A. Rashid, "Real-time COVID-19 diagnosis from X-ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm," *Biomedical Signal Processing and Control*, vol. 68, article 102764, 2021.
- [16] H. Jia, K. Sun, W. Zhang, and X. Leng, "An enhanced chimp optimization algorithm for continuous optimization domains," *Complex & Intelligent Systems*, vol. 8, pp. 1–18, 2021.
- [17] A. Saffari, S. H. Zahiri, M. Khishe, and seyed mohammedreza mosavi, "Design of a fuzzy model of control parameters of chimp algorithm optimization for automatic sonar targets recognition," http://ijmt.iranjournals.ir/article_241126.html.
- [18] M. Kaur, R. Kaur, N. Singh, and G. Dhiman, "SchoA: a newly fusion of sine and cosine with chimp optimization algorithm for HLS of datapaths in digital filters and engineering applications," *Engineering Computations*, vol. 38, no. 2, pp. 975–1003, 2022.
- [19] M. E. Zayed, J. Zhao, W. Li et al., "Predicting the performance of solar dish Stirling power plant using a hybrid random vector functional link/chimp optimization model," *Solar Energy*, vol. 222, pp. 1–17, 2021.
- [20] A. Saffari, M. Khishe, and S.-H. Zahiri, "Fuzzy-ChOA: an improved chimp optimization algorithm for marine mammal classification using artificial neural network," *Analog Integrated Circuits and Signal Processing Signal Processing*, vol. 111, no. 3, pp. 403–417, 2022.
- [21] M. Khishe and M. R. Mosavi, "Chimp optimization algorithm," *Expert Systems with Applications*, vol. 149, article 113338, 2020.
- [22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [23] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *In Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, Perth, WA, Australia, 1995.
- [24] B. S. Yıldız, A. R. Yıldız, N. Pholdee, S. Bureerat, S. M. Sait, and V. Patel, "The Henry gas solubility optimization algorithm for optimum structural design of automobile brake components," *Materials Testing*, vol. 62, no. 3, pp. 261–264, 2020.
- [25] M. Khishe and M. R. Mosavi, "Improved whale trainer for sonar datasets classification using neural network," *Applied Acoustics*, vol. 154, pp. 176–192, 2019.
- [26] M. R. Mosavi, M. Khishe, and M. Akbarisani, "Neural network trained by biogeography-based optimizer with chaos for sonar data set classification," *Wireless Personal Communications*, vol. 95, no. 4, pp. 4623–4642, 2017.
- [27] M. Taghavi and M. Khishe, "A modified grey wolf optimizer by individual best memory and penalty factor for sonar and radar dataset," *Classification*, vol. 6, no. 1, p. 120, 2019.
- [28] M. Khishe, M. R. Mosavi, and A. Moridi, "Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation," *Applied Acoustics*, vol. 137, pp. 121–139, 2018.
- [29] W. Qiao, M. Khishe, and S. Ravakhah, "Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified whale optimization algorithm," *Ocean Engineering*, vol. 219, article 108415, 2021.
- [30] M. Khishe, M. R. Mosavi, and M. Kaveh, "Improved migration models of biogeography-based optimization for sonar dataset classification by using neural network," *Applied Acoustics*, vol. 118, pp. 15–29, 2017.