Al-Rafidain Journal of Engineering Sciences

Journal homepage https://rjes.iq/index.php/rjes

ISSN 3005-3153 (Online)

Improving a Flower Pollination Algorithm to Demonstrate Data Upgrading to Deep Data

Nada Hussain Ali

Computer Science Collage, University of Technology-Iraq

ARTICLE INFO

Article history:

Received	28 January 2026
Revised	05 March 2026
Accepted	12 March 2026
Available online	29 March 2026

Keywords:
data transformation,
FPA,
clustering,
classification,
regression

ABSTRACT

Most computational tasks if not all depends highly on the data formula, as well as the techniques used to achieve these tasks like clustering, classification and regression. Therefor most recent research focus on using data preprocessing as an initial stage. In this research a framework is designed and built that depends on two approaches, the first one is transforming the data into a new formula called deep data through seven structural steps instead of data preprocessing in order to get a data formula that is more robust, reliable and general form. The second approach resides in using the improved flower pollination algorithm (IFPA) relying on local and global search that guaranties the dual behavior to achieve exploitation and exploration behaviors which make the algorithm more flexible in dealing with different tasks like clustering ,classification and regression. The proposed framework is based on obtaining a data formula that is deep data as new solution for processing unstructured and unbalance data in different types textual, numerical and descriptive data types which gave the algorithm an explicit flexibility to manage various tasks. The results shows obvious superiority of the proposed framework in comparison to traditional methods without using Deep data, improved FPA or both, also the proposed framework proved robustness in dealing with clustering, classification and regression as the experimental results showed.

1. Introduction

Data processing is an important task which usually takes a significant percentage of time spent on projects of data science. The time spent on data preprocessing can be caused of the nature of the processing on data which is not a streamlined process. Another issue with data processing is that it’s an iterative approach where the optimal parameters and processing steps can be found by trial and error method. The aim of data processing is to find a data representation for machine learning which yields the best result. [1]

Real-world data is known for its incompleteness and inconsistency which make it crucial for data processing before training and building any model to ensure better performance and reliable results. Data processing is a stepwise and detailed process which starts with data cleaning, integration, transformation and ends up with data reduction data normalization which is considered as the most important part of data processing. [2]

Data normalization is the process of transforming data to a specific range, usually between 0 and 1 or between -1 and +1. When dealing with data that has big differences in the ranges of different features Normalization is required. When the data does not include outliers the scaling method can be considered as useful. Normalization aims to reduce the error of misclassification in classification

Corresponding author E-mail address: Nada.h.ali@uotechnology.edu.iq
https://doi.org/10.61268/8wgzj934

problems and in regression problems can be used to reduce the root mean square error [3].

Data transformation methods remove inconsistencies, aggregate and add data and integrate data. Data transformation commonly refers to the different operations performed on data before it is used to learn models. [4]

Deep data can be defined as rich data that consist of collection of data objects, the deep data may not be large but considered deeper than the raw data [5].

Metaheuristic and nature inspired algorithms has become a popular tool for solving complex optimization problems that are considered as non-linear and multimodality in the field of science and engineering. Recently such tools has been developed, with new nature-inspired algorithms created by academics that could probably be more effective and efficient. One such algorithm is the Flower Pollination Algorithm (FPA). [6]

Flowers pollination algorithm (FPA) is an advanced meta-heuristic algorithm that is considered as an effective intelligent optimization with simple few parameters that have captivated the attention of many researchers in recent years. Nevertheless, the FPA sufferer's from slow convergence, falling in the trap of local optimal, and insufficient accuracy when dealing with problems that are complicated like the operations planning problem [7].

Reproduction is the main purpose of flowers by pollination, depending on pollinators, the process of pollination can be divided into two Categories, biotic and abiotic. Biotic pollination is pollination that involve the organisms while abiotic pollination happens without the involvement of organisms. Pollination can be done in two methods, self-pollination and Cross pollination.

Cross pollination can be interpreted as pollination of flowers in the same tree. Cross pollination can be interpreted as pollination that takes place between flowers that belong to different trees. The transfer of the pollens that are used in the pollination process is usually associated with pollinators, such as birds, insects and other animals. In fact, there is a connection between some flowers and their pollinators. Flowers that have connections with pollinators have the advantage of high percentage of transferring pollens to similar flowers is maximal. Similarly, pollinators benefit from being able to reliably find nectar in similar flowers with minimal learning effort. This behavior can be seen as an optimization process driven by the plants themselves. All the factors and mechanisms involved in flower pollination contribute to maximizing plant reproductive success [8].Several Versions of FPA Algorithm have been successfully applied in several optimization problems and real-world applications such as image segmentation, benchmark optimization functions, feature selection, constrained engineering optimization problems, economic dispatch problems wireless sensor network and many others. The suitable balance between exploration and exploitation forces can be considered as the successful search behavior of FPA Exploitation include an intensive search of solutions that are already existed and have been explored previously, whereas exploration is the ability of the algorithm to search and discover new areas of the search space for further possibilities.[9]

In this paper an improved FPA algorithm is presented to be employed for three tasks, classification, clustering and regression. The aim of the improved FBA algorithm is to test the transformation of raw data into deep data approach proposed by the authors and to test the improved FPA on traditional tasks over the performance of traditional FPA algorithm.

2. Related works

In [10] Mohammed Alweshah et al presented a work aimed to fine tune the weights of neural networks in order to increase the accuracy of classification process to achieve fast convergence. The authors created a hybrid model the used the flower pollination algorithm with Probabilistic Neural Network PNN. The initial solutions where randomly generated by the PNN the the weights of PNN where adjusted by the FPA. Experimental results showed that the hybrid approach performed better than the traditional PNN tested on 11 benchmark datasets and the FPA can get improved results ranging from (80 % to 95 %) in regarding of the classification accuracy.

In [11] Douglas Rodrigues et al evaluated a binary-constrained version of the Flower Pollination Algorithm (FPA) for feature selection process where the search space is presented as Boolean lattice, in this search space each possible solution or string of bits represents if the feature will be used or not to compose a final set of features. Experiment results showed that the FPA outperform other algorithms tested on different datasets in regards of feature selection process.

In [12] Allouani Fouad et al presented a simple modification by using genetic algorithm operators crossover and mutation in order to improve exploitation and exploration abilities, these operators were added after the new candidate solution calculations and the greedy selection, the method proposed by the authors were tested on all the CEC2005 dataset and the experimental results showed that the proposed method were very promising and competitive.

In [13] M. Iqbal Kamboh et al improved the FPA using dynamic switch probability in order to enhance the equilibrium between exploration and exploitation to increase the search ability and diversify the population by using the swap operator. The results of the proposed improvement outperforms standard FPA and other metaheuristic algorithms.

In [14] Muhammad Iqbal et al presented an improved FPA with dynamic switch probability and cross over operator to solve the optimization problems. The proposed method deal with multi constraints optimization problem, because of the few number of parameters the FPA has it is easy to implement, the improved FPA is implemented to solve the complex dispatch of the economic load optimization problem that can define the immature convergence problem by using the cross over operator to enhance and increase the diversity of the population of the local search procedure. The proposed technique were tested against fast evolutionary programming, other versions of fast evolutionary programming, standard FPA and other metaheuristic algorithms using three generator units and thirteen thermal power generation units and by including the effects of valve point loading unit The proposed technique has outperformed other methods in terms of the lowest operating fuel cost

In [15] Zheng, J et al used localization algorithm to enhance flower pollination algorithm (FPA) using Gaussian perturbation and the DV-Hop method. The problem with FPA is the premature convergence and the balance between exploration and local exploitation abilities therefore an improved method is presented. The Gaussian perturbation is introduced to solve the imbalance between the local exploitation and the global exploration in search capabilities, while to exploit the contrast of population information. Based on levy flight and optimal individual an enhanced strategy is proposed, the results showed that the enhanced FPA outperform the standard FPA in terms of convergence and search capability the best value for the normalized mean squared error obtained by the most advanced algorithm, RACS, is 20.2650%, and the best value for the mean distance error is 5.07E+00. However, EFPA-G reached 19.5182% and 4.88E+00, respectively. It is superior to existing algorithms in terms of positioning, accuracy, and robustness.

In [16] WENJING LI et al presented a new adaptive FPA using t-distribution (OTAFPA) and opposition-based learning implemented on social network. The opposition-based learning technique is used to increase diversity and quality of the initial population then the adaptive dynamic switching probability is presented that can effectively balance the local and global search depending on the current number of iterations. At last, the t-distribution variation is utilized to increase the population variety and to aid the algorithm to leap out of the local optimum. The experiments results on eight classical test functions introduced that OTAFPA has better ability on the global optimization, which enhance the convergence speed and the accuracy of the algorithm. The OTAFPA also present superior performance in practical applications of user identification across social networks.

3. Methodology

The Flower Pollination Algorithm (FPA) is a population-based optimization method that operates on a set of candidate or randomly generated solutions. During each iteration, every individual in the population undergoes one of two possible operations: local pollination or global pollination. In the local pollination process, the decision variables of a given solution are influenced by two other randomly selected solutions from the population. In contrast, global pollination guides the current solution toward the best solution found so far. A switching mechanism determines whether the search proceeds locally or globally, and this iterative process continues until a predefined termination condition is satisfied.[17]. FPA abstracts the natural flower pollination mechanism into four idealized rules. **Rule 1** models biotic cross-pollination as a global pollination process, where pollinators move between flowers over long distances. This behavior represents the exploration phase of the algorithm and is typically simulated using Lévy flights to enable a global search of the solution space. **Rule 2** treats abiotic and self-pollination as local pollination, corresponding to the exploitation phase that refines solutions in a limited neighborhood. **Rule 3** represents flower constancy as a reproduction probability, reflecting the tendency of pollinators to favor certain flower types, thereby improving pollination efficiency. **Rule 4** introduces a switch probability, denoted as sp∈[0,1]sp \in [0,1]sp∈[0,1], which controls the balance between local and global pollination. Owing to spatial proximity and environmental factors such as wind, local pollination typically accounts for a significant portion of the overall pollination process. [18]

Flower constancy can be represented mathematically as

$$X_i^{t+1} = X_i^t + \gamma L(\lambda)(X_i^t - B) \qquad (1)$$

Where t i x is the pollen i or solution vector xi at iteration t, and B is the current best solution found among all solutions at the current generation/iteration. While γ is a scaling factor that control the step size. L(λ) is the parameter that represent the strength of the pollination, which can be considered as the step size also. Since insects may travel a long distance with different distance steps, Le'vy flight can be used to imitate this characteristic efficiently, in which L > 0 can be drawn from a Levy distribution:

$$L \sim \frac{\lambda\,\Gamma(\lambda)\sin\lambda\,\pi/2}{\pi}\,\frac{1}{S(1+\lambda)}\ ,\ (S>>S_0>0) \qquad (2)$$

Here, Γ(λ) is the standard gamma function, this distribution can be used for large steps s > 0. Then, to represent the local pollination, both Rule 2 and Rule 3 can be represented as:

$$X_i^{t+1} = X_i^t + U\,(X_j^t - X_k^t) \qquad (3)$$

Where X_j^t and X_k^t are pollen from different flowers of the same plant species. This mechanism mimics flower constancy within a confined neighborhood. Mathematically, when X_j^t and X_k^t belong to the same species or are drawn from the same population, the process is equivalent to a local random walk, assuming U is sampled from a uniform distribution. [19].

3.1 The Proposed Framework

The process of upgrading big data and other data types required special experience and efforts. As it is well known big data requires several processes to deal with it, especially with its characteristics including the 7Vs, Volume which means the magnitude of data, Velocity which means the velocity of data processing signifies the rate at which information is handled and becomes accessible, Variety which means the diversity of data types poses a significant challenge in the realm of big data analytics, Variability, distinct from variety, Veracity, which means Ensuring veracity in data involves maintaining accuracy through processes that prevent the accumulation of inadequate data within systems, Visualization, which is critical in today’s world, it plays a pivotal role in employing charts and graphs to represent extensive and intricate data sets, and Value, which means achieving value stands as the ultimate.

Deep data is a modern and advanced method of data upgrading so the new form ,when it is obtained, will erase or at least reduce some stages of data conversion processes into other formats, including data preprocessing stage (with its various steps) which performs some steps to get a manageable data format, features extraction for reducing complexity of data to transform raw data into smaller and manageable dataset through relevant features and features selection for improving models to remove irrelevant and redundant features from dataset. Deep data is proposed upgrading model aiming to get new data format retains all the properties, formats, specifications, size and variety of data. The basic idea behind presenting this research topic lies in two phases, constructing the deep data model and improving the nature inspired algorithm representing by Flower Pollination Algorithm (FPA). Deep data model includes several organizational processes in order to obtain a formula that makes it easy to deal with any type of data and maintain the privacy and generality of the data at the same time. Seven processes are proposed to deal with deep data model including correlation, relation, structuring, frequenting, homogeneity, generality and stability, through which the previous processes that were performed on the data before the deep data proposal was dispensed with. The proposed deep data model is verified and validated through several tasks performed on different data including classification, clustering and regression task to verify the feasibility of the proposed model. The FPA is a metaheuristic nature inspired optimization method which mimics how plants flowering reproducing during pollination. FPA employs two main processes: cross-pollination, representing the global search for best solution, and self-pollination, representing the process of local search. FPA is characterized by its effectiveness and simplicity to solve various problems types across different tasks, such as classification, clustering and regression.

3.1.1 Turning Big Data into Deep Data: The Seven Steps

To transform big data into deep data, seven particular and practical facility steps are suggested to be applied for getting the upgrading form:

A. **Correlation**: Representing how two or more features are statically correlated and showing the change to each other. This step involves identifying relationships between features to help refine the data for model training. Correlation expressed by coefficients ranging, from -1 which represents perfect negative correlation (the variables move in opposite directions) to +1 which represent perfect positive correlation (the variables move in same direction), while the non-linear correlation indicated by 0. For instance, in the QS Ranking dataset, factors will be correlated like faculty-student ratio with overall university ranking.

B. **Relation**: Data relation is used for relating two features of data table to each other through data column objects. It is an establishment of meaningful relationships between features, making the dataset more structured and tailored to the machine learning task at hand. In the News Category Dataset v3, linked article titles with corresponding categories for classification tasks will be accomplished.

C. **Structuring**: the process of data organizing, storing, and managing for making data easier to access and efficiently process. Structuring involves choosing a specific data format, or structure, which defines the relationships between elements and operations that will be achieved on them. The process of enforced structure in the data is done to get structured form of data from unstructured one, like in the News Category Dataset v3, transforming unstructured text into structured, "machine-readable formats".

D. Frequenting: the frequent of feature is the number of times of the observation has been occurred or recorded in experiment. These frequents are almost depicted in graphic or tabular form. This step involves identifying frequent data patterns and outliers. For instance in the Mall Customer Dataset, this step is used to identify patterns in customer behavior and purchasing habits.

E. Homogeneity: homogeneity represents that the dataset is uniform in nature, thus all its elements will occur with the same type or will share similar characteristics, formats, and/or sources. The homogeneity will allow for simplest analysis and processing, so it will eliminate needing for complex transformations as well as will ensure data integrity. For example, a people dataset will be homogeneous if all elements share the same gender or age. It balancing and normalizing the datasets based on advanced techniques like "RandomOverSampler" to ensure data homogeneity and leading for improving model performance.

F. Generality: Data generality will replace specific values with ranges for enhancing privacy, making data difficult to adversaries for distinguishing individuals within a group depend on their characteristics. Generality will enhance the code’s modularity and making it adaptable to different datasets and/or tasks, by ensuring that it could handle various types of data with minimal modifications.

G. Stability: statistically, it refers to the consistent of data with statistical measure, model over time or/and across different datasets. It is crucial to ensure that the results that obtained from statistical analyses will be reliable and generalize to a broader context. Now it is the time for implementing error handling and reproducibility it by setting random seeds to ensure consistent results across different model runs and prevent variability in performance. figure (1) illustrate the proposed framework flow.

3.1.2. Landscape Analysis for Modified Flower Pollination Algorithm

By incorporating landscape analysis into FPA, this article responds to its shortcomings of applying it in a high-dimensional searching problems environment. The modification of FPA in this article enables the algorithm to change its search strategy depending on the topology of the landscape; which has been shown to result in a substantial increasing in efficiency and convergence rate as well as improved behavior leads to enhanced results. In the light of above description, landscape analysis assists the modified FPA to update its global (which represents exploration search space) and local search (which represents exploitation search space) strategies in call dynamically. The algorithm uses gradient based analysis also estimates population entropy in order to adapt the Lévy flight parameters and switching probabilities. The improved FPA demonstrated superior search behavior and achieved up-to-date results in three different tasks, namely classification, clustering and regression with different datasets.

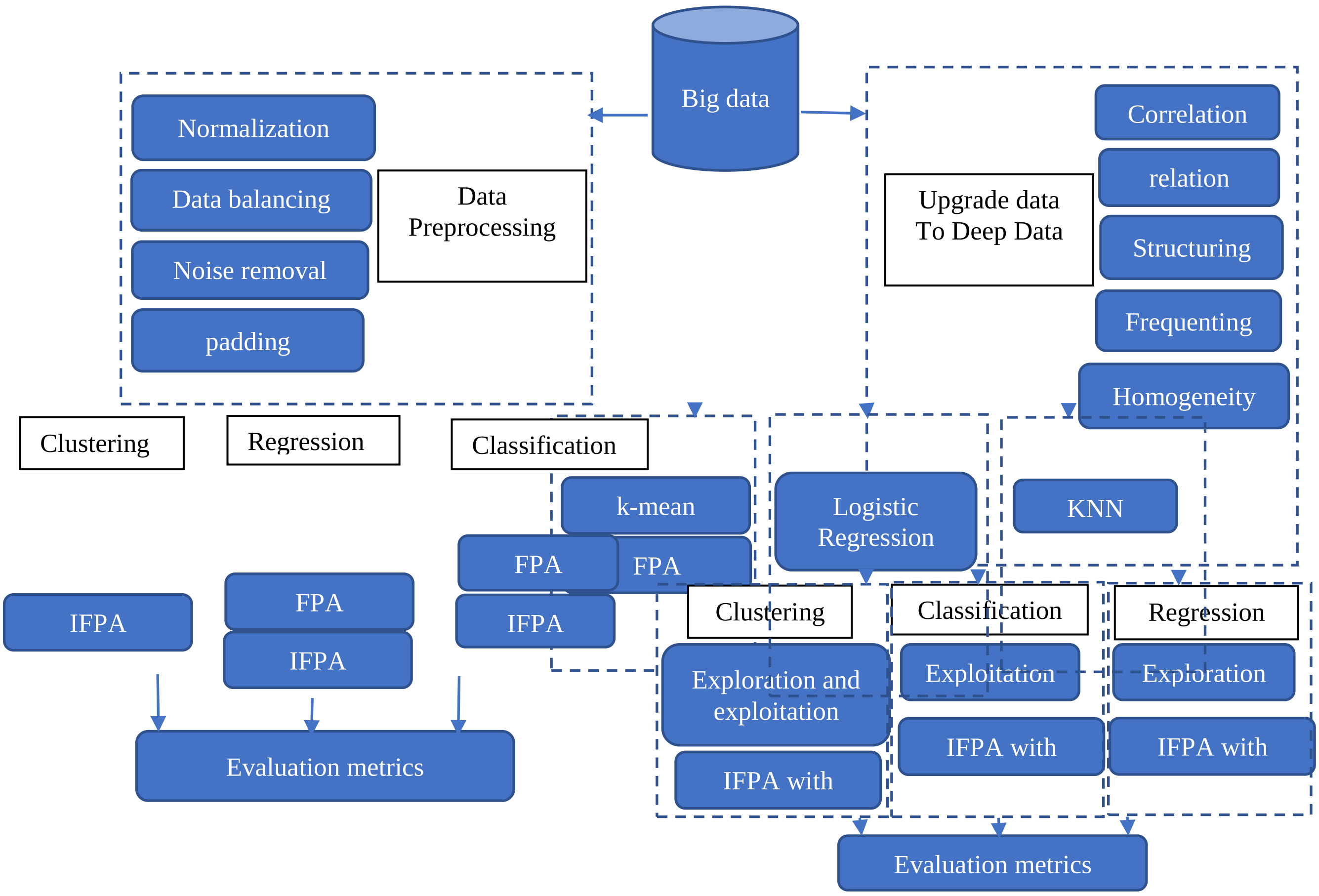

Figure (1), The Proposed Framework flow

Steps of the Modified FPA:

1. Initialization:

- Come up with a selection of flowers (solutions) at random.
- Specify the overall and local Pollination probability at the international and domestic levels (respectively) at the first step.

2. Landscape Analysis:

- Gradient Calculation: Predict the gradient of the fitness space with the numerical differentiation or other methods of differentiation can be used.
- Entropy Measurement: The logistic is to evaluate the entropy of the population in order to determine the amount of diversity and explore-exploit balance.
- Fitness Landscape Topology: it is necessary to analyze the topology of the certain area with the aim to detect the flat regions, ridges, or valleys.

3. Global Pollination with Adaptive Mechanisms:

- Dynamic Lévy Flight: it is therefore important for adapt the Lévy flight parameters with the existing landscape analysis. For instance, with reference to spatial data, one can use increased step size in plains and a reduced step size in mountainous areas regions.
- Entropy-Based Adjustment: low entropy which points to low diversity means that the organization or the society is healthy or in good state. One can also try to explore more

by tweaking the values in the global pollination parameters.

- Local Pollination with Enhanced Local Search: it is gradient-based Local Search, it can be said "let's use the gradient information to fine-tune the local browsing steps", advancing more assertively in the directions that seem most fruitful.
- Entropy Maintenance: local search techniques should be introduced with a view of keeping or improving the standard of the existing products offered by the competitors increase diversity if needed.

4. Adaptive Switch Probability:

- Change the value of the switch probability depends on the characteristics of the landscape. For example, raise probability that benefits world pollination, in regions where the gradient is low.

5. Evaluation:

- Assess the suitability of the new solutions that have been generated.
- Notify the population with improved courses of action.

6. Termination:

- Carry out the process up to when a termination condition is achieved.

4. Datasets Description

Three different datasets are used to evaluate the performance of the proposed framework (Deep data approach with IFPA):

News Category Dataset v3: The HuffPost dataset is one of the largest textual collections that include approximately 210,000 news headlines published from 2012 to 2022. Raw data gives the headlines which are nearly 200000 from 2012 to May 2018 and 10000 from May 2018 to 2022. Every entry in the dataset offers several features; such as the category of the article, its title, list of authors, the source link to the article, a short description or abstract, and the publication date. The dataset is pre-divided into 42 news categories; providing a multitude of avenues for evaluation. It is clear that the most popular categories with the greatest number of articles include politics, wellness, entertainment, and travel in the top 15 categories at our disposal. It was used in the classification problems.[20]

QS Ranking Dataset: The QS World University Ranking is an annual benchmarking survey instituted to assess universities around the world. The current rankings are for 1,500 universities across 104 countries and cover such aspects as employability or sustainability in the 20th edition of QS rankings. In 2024, the ranking introduced significant methodological enhancements, adding three new metrics: including Sustainability, Employment Outcomes for New Graduates and International Research Network as factors that offer a further assessment of universities' societal impact. Comprising 17.5 million academic papers analyzed alongside input from 240,000 academic faculty members and employers, we ensure the comparison has validity and is a sound measure of institutional performance. For the twelfth time in a row, Massachusetts Institute of Technology (MIT) takes the top position, followed by the University of Cambridge and the University of Oxford. This underscores the continued focus of some top institutions in research and teaching and forms for delivering benefits to society. Features from the dataset are various characteristics connected with university rankings, and it was applied to regression tasks for ranking scores prognosis when the given task involves the estimation of the ranking score to a certain University. This dataset contains data including the academic and employer rank and size, student and faculty number.[21]

Mall Customer Dataset: provides data on 200 individuals who visit a mall, including demographic information, annual income, and spending habits. It is customers' data such as gender, age, income level, and the spending score. The available dataset is useful for exploratory data analysis, customer's segmentation, and clustering tasks based on the behaviors of purchasing. It is employed within the clustering context aiming at creating clusters of customers referring to their expenditure.

Mall customer dataset is employed with clustering task, News category dataset v3 is employed with classification while QS ranking dataset is employed with Regression.

5. Results and Discussion

In order to scientifically and practically assess the presented framework of upgrading big data into deep data as well as modified FPA based landscape analysis, the entire implementation process must be carefully considered, the results obtained for each task (clustering, classification and regression) must be evaluated according to the standard criteria for each task.

5.1. Metrics

Three tasks are employed in the proposed framework in order to prove the validity of the deep data approach with improved FPA in solving various problems; these are clustering, classification and regression. The evaluating metrics that are used to assess the results for each task are illustrated as follow:

- **Clustering:** Silhouette Score, Silhouette coefficient, Dunn's Index Calinski-Harabasz Score and Standard deviation.
- **Classification:** Accuracy, Precision, Recall, F1 Score and Error rate.
- **Regression:** Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

5.2. Implementation Results

In this article, comprehensive implementations are performed with the proposed framework (deep data approach and improved FPA) and other techniques (for each task, the most popular technique is implemented) over three datasets used the same stochastic grouping method for clustering, classification, and regression tasks. These techniques are then used on sets to evaluate the kind of enhancements in functionality and results that are obtained. The following results tables show the evaluation statistics for each of the measures defined earlier as well as for each dataset used in the experiment. These outcomes have clear and great impact on the effectiveness of the proposed framework in enhancing the quality of the data (deep data approach), and consequent optimization of the improved FPA used as ML in the proposal, this improvement is due to the fact that adaptive Lévy flight mechanism, which allows for more effective search in the given space.

As a clustering Task, the results of implementing the proposed framework via "Mall Customer dataset" are illustrated in table 1. The experiments are done through five phases, that are, K-mean as popular clustering technique and FPA to be compared with proposed framework phases (the rest three phases) to evaluate the phases of IFPA, deep data approach with FPA, and deep data approach with IFPA

Table 1: Clustering Results with Mall Customer Dataset

Method	Silhouette Score	Silhouette coefficient	Dunn's Index	Calinski-Harabasz Score	Standard deviation
K-Means with processed data	0.738	0.728	0.749	0.817	0.281
FPA with processed data	0.712	0.692	0.702	0.786	0.290
IFPA with processed data	0.763	0.751	0.766	0.833	0.278
FPA with deep data	0.856	0.823	0.848	0.887	0.232
IFPA with deep data	0.873	0.844	0.869	0.911	0.224

Regarding the use of data in its known format, the results show that K-mean slightly superior to FPA, but the IFPA outperforms both of them in term of widely dependent clustering evaluating metrics. As for proposal's point of view and from the perspective of the proposal through the use of the data driven, deep data approach, the results show the effectiveness of data upgrading which clearly demonstrates the superiority of the proposed deep data approach without the improvement FPA and even greater superiority with IFPA. It is worth noting here that two essential perspectives are embedded, firstly, the proposed framework provides the knowledge with a clear application of exploitation and exploration, represented by the presence of clusters and their elements, secondly, the role of landscape analysis in enabling the IFPA to unleash its potential as a noteworthy clustering technique.

The proposed framework is employed to accomplish the classification task using the "New **Category Dataset v3**", table 2 clarify the noticeable enhancement of the results according to the five phases that mentioned in the analysis and discussion of table 1.

Table 2: Classification Results with News Category Dataset v3

Method	Accuracy	Precision	Recall	F1-score	Error rate
KNN with processed data	82.02	83.91	86.14	85.01	17.98
FPA with processed data	81.29	83.46	85.43	84.43	18.71
IFPA with processed data	85.08	85.82	86.42	86.12	14.92
FPA with deep data	88.41	88.33	92.09	90.17	11.59
IFPA with deep data	96.05	96.98	96.66	96.82	3.95

It is clear that IFPA achieves a superior classification rate than KNN and FPA via processed data; however, KNN is the closest to FPA due to their employment of processed data under the same classification conditions based exploitation application. When applying FPA to the upgrading form of data, the upward trend become evident in obtained results, and even more with the IFPA, also the impact of deep data approach and landscape analysis are obvious in exploitation process.

The third task that assigned in the proposed framework is the regression task by using the "QS Ranking Dataset" based exploration application, table 3 demonstrates the error rate depending on famous and popular error criteria as illustrated bellow.

Table 3: Regression Results with QS Ranking Dataset

Method	MAE	MSE	RMSE	R^2
LR with processed data	0.376	0.327	0.243	0.803
FPA with processed data	0.391	0.342	0.266	0.739
IFPA with processed data	0.334	0.283	0.247	0.816
FPA with deep data	0.223	0.226	0.218	0.857
IFPA with deep data	0.194	0.207	0.202	0.874

From table 3, logistic regression as a popular and standard regression technique is employed as well as the FPA to be states of the art, the first three experiment phases (LR, FPA and IFPA) over processed data showed similar error rates results, with a clear advantages for IFPA due to the same common reasons. The presence of the new data form also indicates a high positive impact on the results of error rates (MAE, MSE, RMSE and R^2), the effectiveness of deep data approach based FPA, as in fourth experiment phase, while the highest error rates results appeared in the fifth experiment phase through the integration between deep data approach and IFPA based landscape analysis.

According to the above three tables results with different tasks, the proposed framework achieves several promising accomplishments, generality in processing various dataset types, robustness against different tasks, reliability of results and search in local and global search space to achieve exploitation and/or exploration.

6. Conclusions

The findings of this research are in tandem with the hypothesis that the IFPA readily harnesses big data, achieving deep data transformation and substantial quality enhancement of large data corpus enhancing consequent machine learning. These findings also support the generality and validity of the proposed framework of deep data approach and IFPA further and offer considerable support for the notion that the movement from big data to deep data is the key for achieving higher levels of machine learning model performance. It can be concluded that the deep data approach reflects a significant advance over the present methods for data processing and has the potential to become widely used in a number of disciplines where high standard of data is important for obtaining accurate and precise outcomes. Ultimately, the results prove this hypothesis correct, as the FPA refines disparate and bulk data to work better for the ML processes. When benchmarked by selecting performance indicators, the framework achieves better outstanding performance across various categories of machine learning models. More particularly, the gross obtains clustering performance is almost fully charged, which showed that the framework accrues good results in grouping and categorizing the data. Similar to binary classification tasks the accuracy is also found to be high which points to the potential of the IFPA in enhancing the results of tasks. The observed results of the framework in regression tasks support the hypothesis of a direct relation between the refinement of the quality of deep data and the subsequent increase in the effectiveness of using the developed framework in various ML fields.

References

[1] Sebastian Strasser, Meike Klettke, Transparent Data Preprocessing for Machine Learning. *In Workshop on Human-In-the-Loop Data Analytics (HILDA 24)*, Santiago, AA, Chile. ACM, New York, NY, USA, 6 pages, June 14, 2024. https://doi.org/10.1145/3665939.3665960

[2] V.Sathya Durga, Thangakumar Jeyaprakash**,** Data Transformation 10 Techniques for Academic Datasets, *International Journal of Engineering and Advanced Technology (IJEAT)***,** Volume-9 Issue-1, October 2019.

[3] Peshawa Jamal Muhammad Ali, Rezhna Hassan Faraj, Data Normalization and Standardization: A Technical Report, *Machine Learning Technical Reports*, 2014.

[4] Agata Kozina, Data transformation review in deep learning, *CEUR Workshop Proceedings*,2024.

[5] Marcin Szczukaú, Dominik ´Sl¸ezak, How Deep Data Becomes Big Data, *Conference Paper Interdisciplinary System for Interactive Scientific and Scientific-Technical Information" founded by Polish National Centre for Research and Development (NCBiR)·* June 2013.

[6] Sahil Lalljith, Ismail Fleming, Umeshan Pillay, and et. al, Applications of Flower Pollination Algorithm in Electrical Power Systems: A Review, *IEEE access*, VOLUME 10, 2022.

[7] Thi-Kien Dao, Trong-The Nguyen, Vinh-Tiep Nguyen and et. al**,** Hybridized Flower Pollination Algorithm and Its Application on Microgrid Operations Planning. *Applied Science*, 2022. https://doi.org/10.3390/app12136487

[8] Derby Prayogo Samdean, Herry Suprajitno, Edi Winarko, Flower Pollination Algorithm (FPA) to Solve Quadratic Assignment Problem (QAP), *Contemporary Mathematics and Applications* Vol. 1, No. 2, 2019, pp. 121-130.

[9] Moh'd Khaled Yousef Shambour, Ahmed A. Abusnaina, Ahmed I. Alsalibi, Modified Global Flower Pollination Algorithm and its Application for Optimization Problems, *Interdisciplinary Sciences: Computational Life Sciences*, 2019,Pp:496–507 https://doi.org/10.1007/s12539-018-0295-2

[10] Mohammed Alweshah, Moad Abu Qadoura, Abdelaziz I. Hammouri and et. al, Flower Pollination Algorithm for Solving Classification Problems, *Int. J. Advance Soft Compu. Appl, Vol. 12, No. 1, March 2020*

[11] Douglas Rodrigues, Xin-She Yang, André Nunes de Souza and et. al, Binary Flower Pollination Algorithm and Its Application to Feature Selection, *Springer International Publishing Switzerland* 2015.

[12] Allouani Fouad, Kai Zenger, Xiao-Zhi Gao, A Novel Flower Pollination Algorithm based on Genetic Algorithm Operators, *EUROSIM 2016 & SIMS 2016, Proceedings of the 9th EUROSIM & the 57th SIMS September 12th-16th Oulu Finland*, 2016. DOI:10.3384/ecp171421060

[13] M. Iqbal Kamboh, Nazri Bin Mohd Nawi, Azizul Azhar Ramli and et. al, An Improved Flower Pollination Algorithm for Global and Local Optimization, *International Journal On Informatics Visualization***,** *JOIV : Int. J. Inform. Visualization*, Pp:461-468, December 2021.

[14] Muhammad Iqbal, Nazri Mohd Nawi, Radiah Bt. Mohamad, An improved flower pollination solution for economic dispatch with valve point effect, *Indonesian Journal of Electrical Engineering and Computer Science* Vol. 22, No. 2,pp. 629~637, May 2021.

[15] Zheng, J.,Yuan, T., Xie, W. and et. al, An Enhanced Flower Pollination Algorithm with Gaussian Perturbation for Node Location of a WSN. *Sensors*, 2023. DOI: https://doi.org/10.3390/s23146463

[16] Wenjing Li , Zhiming He, Jian Zheng And et. al, Improved Flower Pollination Algorithm And Its Application In User Identification Across Social Networks, *Special Section On AI-Driven Big Data Processing: Theory, Methodology, And Applications*, Volume 7, 2019. DOI: 10.1109/Access.2018.2889801

[17] Dr. BN Shobha, Soujanya B Sundaresh, Flower Pollination Algorithm, *International Research Journal of Modernization in Engineering Technology and Science* Volume:04, Issue:07,July-2022.

[18] Safaa Saber, Ibrahim Elhenawy, A Survey on Flower pollination algorithm, *Journal of Intelligent Systems and Internet of Things*, Vol. 2, No. 1, PP. 05-11, 2020.

[19] Nabil Diab, Emad El-Sharkawy, Recent Advances in Flower Pollination Algorithm, *International*

Journal of Computer Applications Technology and Research, Volume 5– Issue 6, Pp: 338 - 346, 2016.
[20] Misra, Rishabh, News Category Dataset,2022.
[21] Joakim Arvidsson, QS World University Rankings,2024
[22] Muhammad Yasir Saleem, Mall Costumer Dataset, 2024.