

**ENHANCED SCHIZOPHRENIA PREDICTION USING HARMONY
SEARCH ALGORITHM BASED LIGHT GBM ML METHOD**

**¹Dr. D. Sasikala, ²Dr. K. Venkatesh Sharma, ³Sathya Selvaraj Sinnasamy,
⁴Dr. M. Pushpavalli**

¹Professor, NIMS University, Rajasthan, Orcid: 0000-0003-4532-575X

²Professor, CVR College of Engineering, Hyderabad, Orcid: 0000-0003-4333-0491

³Assistant Professor, Department of Electronics and Communication Engineering

SRM institute of Science and Technology - Ramapuram

Chennai, Tamil Nadu, India , Orcid: 0000-0003-3676-0178

⁴Associate Professor, Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu 638401, India. Orcid: 0000-0001-9329-3671

¹godnnature@gmail.com, ²venkateshsharma.cse@gmail.com, ³sathyas5@srmist.edu.in,
⁴pushpavallim@bitsathy.ac.in

Abstract

Schizophrenia (SCZ) discovery is holding result on various investigations and for patients pleasant lives swift and clear-cut dedication is important. Scans from multifarious modal gadgets specify contradictory effects of schizophrenia. Schizophrenia labels brought about on the achieved factual data is by numerous errands of the ML systems. These are brought together concurrently allowing the succeeding details – (1) the ML system ascertained that was applied. (2) The estimates illustrating the traits in the process and (3) the system by which introductory data was secured. The Light gradient boosting (LGBost) classifier is procured separately to categorize the three subtypes of SCZ by estimating the magnetic resonance imaging (MRI) data, instead the Amplitude of the Low-Frequency Fluctuations as well as the Gray Matter Volume with their acronyms as ALFF & GMV are the aspects of classifiers. Later, the Dempster–Shafer (DS) Theory of evidence is set in to get blend to reveal the substantial likelihood errands imparted on the returns of inconsistent categorizers. The three labeling are deficit schizophrenia (DS), nondeficit schizophrenia (NDS) on or after healthy control (HC) that be sure to ascribe to entirely put in storage on the field in study. Accordingly, investigation indicates that a hyper-parameter optimization scheme that brings in greater attainment in lesser time than the contemporary Grid search approach is intervened with the Harmony Search (HS) to the win through LGBost system with SZ analysis for sub-types study is implied. A coherent review exposes that the anticipated HSLGBM hit the highest point in the leading-edge approaches. Authenticated on this data, HSLGBM is proficient in forecasting the prospects of actuality intended for either DS, NDS or HC with 90.71% accuracy. On the other hand this is still to be enhanced with innovative methodologies so that SCZ forecast prefer to be gently superior in the imminent investigation works.

Index Terms: Multi-class Classification ML algorithm – Light Gradient Boosting Method, GMV, ALFF, DS, NDS, healthy control (HC) and Meta-heuristic Optimization Harmony search algorithm.

1. INTRODUCTION

Schizophrenia (SCZ) syndrome impacts bring about concerns in entire parts of life concerning the persons besides their family, educational, societal, and professional functioning. If a group of efficacious caution occasions are dedicated, no hardly any one in three individuals will be proficient of retrieval fully with schizophrenia [1]. It is thought about as an inveterate psychosomatic health syndrome that survives on a band that embraces Simple, Disorganized/Hebephrenic, Paranoid, Catatonic, Residual, Undifferentiated, Cenesthopathic, Unspecified

schizophrenia, together with schizoaffective, brief psychotic, delusional, schizophreniform, schizoaffective disorder and much more. For that reason it's not extensively detected with subtypes. [2]. Schizophrenia scrutiny is induced by a sum of assessments and for those sociable lives early and exact immersion is vibrant [3]. CT and EEG spectral analysis, DTI, MRI, and PET images indicate dissimilar results of schizophrenia. ML approaches is handled to recognize SZ by scrutinizing exhibits in data with certain suggestive conditions. These data are initiated and stored from many sources, for illustration from person records, brain imaging images, or so far from social media set ups. So its swift analysis is presently required so its therapy will be speedier [4].

ML for medical review facilitates detecting SZ illnesses in the preliminary stages. Schizophrenia categorization set up on the procured specific data is by plentiful purposes of the ML approaches. These are accumulated as one considering the subsequent details – (1) the ML system ascertained that was applied. (2) the estimates illustrating the traits in the process and (3) the system by which introductory data was secured. KNN, AB, LDA, LR, SVM, DT, RF, MLP, ANN, Feed forward Neural network, CNN, DNN, XGBoost, LGBost and much more processes were applied for Schizophrenia forecasting [5]. This work stresses on the categorization of ill individuals with schizophrenia through an already available XGBoost (eXtreme Gradient Boosting) scheme and a presently implemented LGB (Light Gradient Boosting) scheme. A supervised and a replacement of gradient boosting, is held for its potential in augmenting swiftness and function superiorly instead of XGBoost as the LGBost ML pattern. SOCR Data Oct2009 ID NI and open public dataset was surveyed in this exploration. The LGBost categorizer is taken as one to discerned as one of the three subtypes of SZ by evaluating the magnetic resonance imaging -MRI data, while ALFF as well as GMV are extracted as the attributes of categorizers. Later, the Dempster-Shafer (DS) theory of evidence is utilized to execute the blend to reveal the crucial probability concerns made from the gains of contradictory categorizers. The DS, NDS or HC were assigned to all solo saved area in study [6].

Survey the blend of hyperparameters that yields excellent performance and an wide category of search schemes – Manual Search, Random Search, Grid Search, Harmony Search and Bayesian Search are executed. To avoid ineffectiveness due to time draining scheme, it is vital to evaluate the technique to detect an optimized blend of hyperparameters conveying consistent depiction in less time. As a result, in many optimization procedures, the harmony search (HS) is fixed by a simple form in fusion of LGBost presently than the old procedures, that includes the particle swarm optimization (PSO) or genetic algorithm (GA) or HS proficiently striving for the results. Thus, in this exploration, a hyperparameter optimization approximate that returns coherent execution in a reduced amount of time than the present Grid search scheme is by the HS to the powerful LGBost tactic with SZ analysis for form subtypes discovery is inferred [7]. A rational exploration shows that the suggested process crowns the existing methodologies.

2. RELATED WORKS

A proficient methodology for future categorization, initial analysis, and detecting of Schizophrenia using LightGBM without and with segmentation [8]. The outcomes put forward the capability of LightGBM as a consistent means for precisely classifying psychological disease-linked subreddits [9]. The fusion system provides high accuracy than the GMV or ALFF categorizer using the small datasets. The augmented multiclassification system's accuracy rate that is put up on Xgboost and fusion technique

is better than other ML methods that is capable of contributing beyond in the review of the clinical schizophrenia [6]. The intended LightFD classifier has improved execution in actual EEG psychological circumstances forecast, that is meant to have a massive use of estimates in brain-computer interaction or shortly as BCI. [10]. The outcomes prove that the LightGBM-made approach overtakes present approaches, realizing viable implementation in the precise detection of persons with schizophrenia [11].

HS had revealed hopeful application in supporting multifaceted fine-tuning challenges and thus numerous adaptations of this system were defined. In the imminent future, HS is engaged to extra real-time optimization disputes [12]. The experimental outputs clearly signify that Integrative Hope Scale (IHS) procedure will efficiently afford enhanced outcomes for solving the utility grid setback evaluated with other fine-tuning systems [13]. The optimally parametrized traditional UFLS approach shown in this manuscript exposes excellent outputs opposed to the traditional UFLS methods, and the fitness of utilizing the IHS procedure is validated [14]. The Differential HS (DHS) was located to overtake the classical HS and its two options such as GHS and HIS with regard to finishing accuracy, unification speed and robustness [15].

3. DESCRIPTION OF THE PREVAILING WORKS

The aim of this exploration work is to propose and create a ML-built optimized means that is responsible for tools, where decision-makers will construct, compute and investigate the essential attributes associated to SCZ detection and assessment. Various measures and evidence have to be considered to offer a ML approach that calculate approximately and provides interpretations to lead an apt prediction in SCZ discovery. In addition to that an inherent and accessible interface has been built-up to provide a signal of the SCZ being explored and its predictions. Besides, this device has been united with Harmony search algorithm (HSA) to deliver a proactive exact analytical scheme built on the SCZ investigation [4, 5, 6, 7, 8, 9, 10, 11, and 12].

An evaluation of the measures of the ML model is also provided. Formerly, it suggests an assessment of the ML-built tool for SCZ calculation and affords an execution comparison of the advanced optimized ML approach as well as offered with numerous considerations about potential encounters / restrictions of the current patterns – LGBost and optimization by HS [4, 5, 6, 7, 8, 9, 10, 11, 12]. But, it can be evidently witnessed by the entities near the diagonal that the LGBost recorded to some extent lower than HSLGBM. Lastly, sum up and completes the document and offers a stance for future work to advance the result. *HSLGBM*: Intended and applied for HS optimization built LGB ML technique that have been trained, tuned by Metaheuristic HS process and exhaustively verified.

4. DEPICTION OF THE PROPOSED WORK

The HSLGBM Approach

HSLGBM: Planned and applied an Ensemble delivered gradient-boosted decision tree ML procedure LGBost that was optimized with HS that was trained, fine-tuned and exhaustively tested. This discovery intends a system for leading qualitative SCZ evaluation. In this set up, the outcomes are initiated, opening with an elite summary of its HSLGBM system architecture with a description of all the modules involved. After that, the ML-built SCZ evaluation workflow is depicted and the prospects of every stage is openly outlined. In

conclusion, information on the unification with the HS's was delivered. An architecture in illustration 1 emphasizes the system components' and their collaborations. Basically the clients are adepting to log on the dashboard through any browser even not including an account. SCZ-linked KPIs, enhance categorization rate increasing the forecasts allied to the MRI SCZ estimation of vital properties. In addition the client is capable of replacing both the appropriate facts and the other factors essential for the SCZ estimation done over the web-based interface.

To make an tangible SCZ forecast, a call for the *SCZ Categorizier* is forwarded. The SCZ Categorizier is a forecast service built-in in the ML *Categorizier* Layer and is actually manipulated to open up the trained ML simulations. The activity of training, substantiating and examining the ML simulations demands abode, then provide with adequate level of details. To end the interface shown gives a supervising stage of the deployed prototypes saving simulation-fact metadata (e.g., accuracy) and many metrics on the forecast service.

SCZ Estimation Venture : Using LGBost ML to SCZ estimation, by constructing and evolving a HSLGBost ML venture (Figure 1) is displayed.

A flow diagram of the supervised ML venture realized in this research, as an important quantity of the result in Figure 2.

Data Gathering: Data is formed by formulae that reflect the numerical assets of the original data not revealing any authentic fact about the topics. By in-depth investigation and review of SCZs and its resultant related facts, the subsequent GMV and ALFF parameters worked as core of this investigation that were seen:

56 area cortical and subcortical volumes (column heading) along with all their parameters:

$$ALFF(x) = \sqrt{\sigma^2(x)} \quad (1)$$

$$\sigma^2(x) = \frac{1}{N} \sum_t (h(t) * S(x, t) - \mu(x))^2 \quad (2)$$

$$\mu(x) = \frac{1}{N} \sum_t (h(t) * S(x, t)) \quad (3)$$

The latest value of the qualitative SCZ assessment put using the past given parameters. As the data launch practice is meant to cause the individuals historical records comparable, the value of the SCZ column will be derived from the past formal or tailored qualitative SCZ estimation models will have one of the subsequent values: "DS", "NDS" or "HC", where manual labelling is very costly taht takes in jillion data.

In directive to produce the facts pointed out over, certain notions that were rendered. Initially, the upper/lower limits for all columns were stated, so that all created value is well defined within the range. The Table shows a framework of the fixed limits for all created facts in this work [6].

Normalization done to scale data on a limit of [0, 1]. Few ML simulations are greatly subtle to properties with unreliable degrees of amplitude, limit and factors. The dataset for this work add in properties for example "GMV", "ALFF" and blend that have unique ranges and training complex simulations on unscaled data that heads to drop in their functioning level and measure- accuracy.

HSLGBM

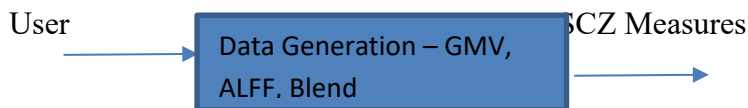


Figure 1. Architecture

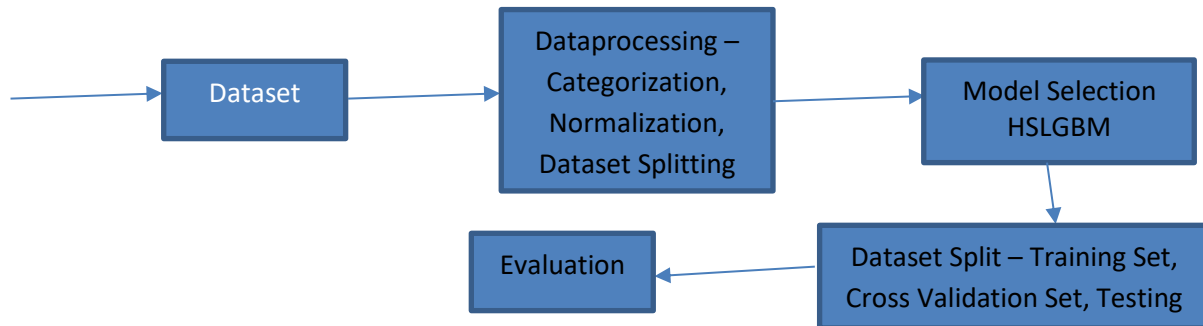


Figure 2. ML Venture with HSLGBM

Here, a Min-Max scaling normalization process known is stated by the resulting formula:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

This is put in to the full dataset especially to “features”, of every single column apart from “SCZ” that have the three outcome categories built on the future forecasts that will be made. Illustration 1, is the last step in this stage dividing the dataset into a set as training, substantiation and examination.

Multi-Class Classification Algorithms

The ML Multi-Class Categorization simulations (MCC) resolves complications of categorizing illustrations into one of three or more outcome categories. In the simulation picking stage - Figure. 2 trendy MMC prototypes are picked for handling qualitative SCZ estimations. The core aim is to propose and acquire ML simulations that are based on realizing related facts to do precise qualitative SCZ estimation forecasts, also extend supervision of the set up beyond looking after continuous estimation built on input data.

Supervised Learning (SL) Light Gradient Boosting Method (LGBM) is a prototype for Multi-class classification and Meta-heuristic optimization Harmony Search Algorithm (HSA) is an approach that are available in the prevailing outcomes.

Unification of these two prevailing algorithms are fused into a new recommended model for classification used as the proposed solution.

Proposed HS + LGBM algorithms

LGBM departs the tree leaf-wise set against other boosting approaches that grow up the tree level-wise. Leaf choices done based on the utmost delta deficit growth. Since the static leaf is available, its approach has a less significant loss associated with the level-wise method.

Gradient boosting structure, LightGBM serves a tree based learning envisioned to be scattered and proficient with the subsequent benefits:

Swift training speed and superior efficiency, reduced memory handling, with high accuracy, maintain parallel, dispersed, and GPU learning and adopting large-scale data.

LGBBoost Benefits

1. Swifter training speed and greater proficiency: Light GBM utilize histogram built approach i.e it buckets incessant aspect amounts into discrete bins that affix the training process.
2. Lesser memory use: Substitutes nonstop amounts to distinct bins that effect in reduced memory use.
3. Enhanced accuracy among the other boosting approaches: It yields considerably further multifaceted trees by following the leaf wise rather than a level-wise splits that is the main property for obtaining superior accuracy. Setting the max_depth parameter will avoid facing the rarely head overfitting.
4. Attuned by Huge Datasets: It is proficient in executing equivalently noble huge datasets with an important drop in training time as related to the LGBBOOST.
5. Backed by using Parallel learning.

Algorithm for LGBM:

Step 1: Data Pre-processing

Step 2: Splitting the data

Step 3: Initialization

Step 4: Splitting the nodes

Step 5: Building the tree

Step 6: Gradient Boosting

Step 7: Regularization

Step 8: Prediction

Step 9: Model Evaluation

Step 10: Parameter Tuning

HSA: A meta-heuristic populace-based search approach that is motivated by the fact that musicians frequently fine-tune the pitch of their gadgets to realize a stunning harmonic state. It generally has a reduced quantity of amendable parameters that is simple to execute.

The benefits of HS primarily embrace: (1) Decision variables is not set, (2) Limited control parameters are essential for fine-tuning, (3) Derived facts is not necessary. (4) Easy to understand. (5) Depict the step-wise result to a known setback. (6) Hindrances are cracked down into smaller pieces or steps, so that, it is at ease for the computer programmer to alter it into a real program. (7) Aid finding results to setbacks swiftly than the traditional methods, and will do so with least data. (8) Find answers to complications that are challenging to verbalize scientifically.

Algorithm for HS Technique:

Step 1: Start the optimization setback choosing its parameters. Index the objective function $f(x)$. Pick the judgement variables and their limitations.

Explanation of the procedure's parameters: Harmony Memory Size, Harmony Memory Consideration Rate, Pitch Adjusting Rate, also Maximum number of iterations indicating the

HMSize, HMCR, PAR and maxIter respectively.

Step 2: Defining the Harmony Memory Matrix Production using the random numbers that do not interrupt the limitations that are set for the judgement variables that supply values of the Harmony Memory Matrix.

Ordering of the Harmony Memory fitting to the execution of every single harmony.

Step 3: Creating a fresh harmony by the three ways: Harmony Memory customization for (a)HMCR (b) PAR and (c) Progress

Step 3.1: Is the ending condition achieved? If Yes then End or No then Goto Step 3.

Step 3.2: Is the innovative harmony superior than the worst put in storage in the Harmony Memory? If it is Yes then Goto Step 4 or No then Goto Step 3.1.

Step 4: Harmony Memory is refreshed to Goto Step 3.1.

Architecture :

LGBM MRI data (GVM, ALFF & Blend) → HSA → DS, NDS, HC

5. EVALUATION OF RESULTS

Performance metrics: A pool of data that firms estimate next to a customary objective.

Table 1. LGBM & HSLGBM Measures

ML Model	LGBoost (GMV)	LGBoost (ALFF)	LGBoost (Blend)	LGBoost + HS
Accuracy	68.36%	72.11%	76.26%	90.71%

Table 2. Computed LGBM & HSLGBM

ML Model	LGBoost (GMV)	LGBoost (ALFF)	LGBoost (Blend)	LGBoost + HS
Precision	65.73%	67.25%	72.18%	81.69%
Recall	69.32%	72.59%	79.71%	91.22%
F1-Score	0.6182	0.6527	0.7119	0.7934
AUC	0.8128	0.8927	0.9261	0.9582

Furthermore, the evaluation of the LGBM ML algorithm is presented and demonstrated that for a dataset using Python programming with its libraries – Numpy, Tensorflow and Pytorch, HSLGBM achieve a slightly higher accuracy. Nonetheless, the two sampling approaches for learning from imbalanced ML methods datasets. Executed favourably mostly adept accuracy accomplishing 90.71%. Besides, the created measures it also approves again that the estimated ML procedures were up to categorizing utmost samples precisely. All the other vital metrics, also specified valuable understandings in the ML approach operation for every single outcome category.

Presently, the leading restriction is the absence of real-life datasets for training the ML approaches exercised in this research work. To rise above partly from this drawback showing

the success of the simulation, a prefixed dataset was measured. But, when this data have many properties and facets that are generally very tough to create high-quality data for multifaceted setbacks. If the produced dataset doesn't pair with its counter parts, the behaviour and properties of the real-life dataset, that will undesirably effect the execution of the trained ML simulations.

To end, LGBM is operated to guess the SCZ of only two types as DS or NDS from HC. To tackle this drawback of categorizing accuracy, the recent prototype was proposed to be clearly extensible innovative approach LGBM was unified with HSA trained precisely for categorizing new types of SCZs i.e., simply be combined into this current outcome.

6. Conclusions & Future Scope

To construct and execute HSLGBM, a GB ML-based HSA optimized tool for reinforcing the SCZ assessment process. Based on this reason, one dataset was obtained and employed in the training and examining stage of every other LGB ML approach. As soon as examined and substantiated, this categorizer is merged with HSA that showed an advanced classifier largely utilized for SCZ diagnosis.

The implemented system estimate the SCZ only for MRI scans, namely GMV, ALFF and their blend. For this, the model needs a precise set of features, and known as profile or related facts. The HSLGBM is able to forecast the likelihood of being committed for DS or NDS from HC with 90.71% accuracy that is to be progressed so that SCZ forecast will be even more better.

Future enhancements take in exploring and looking into aspects that give emergent SCZ-precise ML approaches, i.e., they adept and fully dedicated to estimate the definite types of SCZs, letting more widespread review phase. Additionally, the experimental examinations are required to estimate the activities and executions of the ML approaches now realized using HSLGBM on public real-life datasets. Moreover, the option of saving forecast feedback will be explored in future studies to find if the functioning and overall accuracy of the ML approaches will benefit from it. This will also test the properties of non stop SCZ supervising, where key performance indicators (KPIs) are frequently obtained using it for automated and non-stop SCZ reviews in order to estimate the probability of unpredictable SCZ accountabilities. Then, this current simulations too aids the integration with a recommended dealing facilities built on the profile also influenced by the calculated SCZ that is to be executed as the future work in this proposed HSLGBM based SCZ diagnosis and analysis system.

References:

1. Laursen TM, Nordentoft M, Mortensen PB. Excess early mortality in schizophrenia. *Annual Review of Clinical Psychology*, 2014;10, 425-438.
2. Heather Jones, "What Are the Different Types of Schizophrenia? - Subtypes No Longer Used in Diagnosis", Article in verywellhealth, November 27, 2023.
3. WebMD Editorial Contributors, "Schizophrenia Diagnosis & Tests: How Doctors Know If Someone Has It. ", Article in WebMD, January 12, 2023.
4. Ngumimi Karen Iyortsuun, Soo-Hyung Kim, Min Jhon, Hyung-Jeong Yang, and Sudarshan Pant, A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis, *Healthcare (Basel)*, 2023 Jan 17;11(3):285. doi: 10.3390/healthcare11030285.
5. Shradha Verma, Tripti Goel, M Tanveer, Weiping Ding, Rahul Sharma and R Murugan, Machine learning techniques for the Schizophrenia diagnosis: A comprehensive review and

future research directions, Springer Nature 2022, pp 1-15.

6. Wenjing Zhu, Shoufeng Shen, and Zhijun Zhang, Improved Multiclassification of Schizophrenia Based on Xgboost and Information Fusion for Small Datasets, Hindawi Computational and Mathematical Methods in Medicine, 2022, vol. 2022, no. 1581958, pp 1-11, <https://doi.org/10.1155/2022/1581958>.
7. Seong-Hoon Kim, Zong Woo Geem and Gi-Tae Han, Hyperparameter Optimization Method Based on Harmony Search Algorithm to Improve Performance of 1D CNN Human Respiration Pattern Recognition System, MDPI, Sensors 2020, vol. 20, no. 3697, pp 1-19, doi:10.3390/s20133697.
8. Pradnya Rajendra Jadhav and Raviprasad Aduri, ChronoPsychosis: Temporal Segmentation and Its Impact on Schizophrenia Classification Using Motor Activity Data, arXiv:2311.12590v1 [cs.LG] 21 Nov 2023.
9. Kannan Nova, Machine Learning Approaches for Automated Mental Disorder Classification based on Social Media Textual Data, JIBSS, Contemporary Issues in Behavioral and Social Sciences, vol 7, no. 1, 2023, pp 70-83.
10. Hong Zeng, Chen Yang, Hua Zhang, Zhenhua Wu, Jiaming Zhang, Guojun Dai, Fabio Babiloni, and Wanzeng Kong, A LightGBM-Based EEG Analysis Method for Driver Mental States Classification, Hindawi Computational Intelligence and Neuroscience, vol. 2019, no. 3761203, pp 1-11, <https://doi.org/10.1155/2019/3761203>.
11. Elizabeth Mayorga Aldaz, Roberto Aguilar Berrezueta, and Neyda Hernandez Bandera, An Intelligent Schizophrenia Detection based on the Fusion of Multivariate Electroencephalography Signals, Fusion: Practice and Applications (FPA), vol. 13, no. 02. pp. 42-51, 2023.
12. Alireza Askarzadeh and Esmat Rashedi, Harmony Search Algorithm: Basic Concepts and Engineering Applications, pp 1-36, 2017, DOI: 10.4018/978-1-5225-2322-2.ch001.
13. Wei Sun and Xingyan Chang, An Improved Harmony Search Algorithm for Power Distribution Network Planning, Hindawi Publishing Corporation, Journal of Electrical and Computer Engineering, vol. 2015, no. 753712, pp 1-6, 2015, <http://dx.doi.org/10.1155/2015/753712>.
14. Jose Miguel Riquelme-Dominguez, Martha N. Acosta, Francisco Gonzalez-Longatt, Manuel A. Andrade, Ernesto Vazquez, and Jos'e Luis Rueda, Improved Harmony Search Algorithm to Compute the Underfrequency Load Shedding Parameters, Hindawi International Transactions on Electrical Energy Systems, vol. 2022, no. 5381457, pp. 1-15, <https://doi.org/10.1155/2022/5381457>.
15. Prithwish Chakraborty, Gourab Ghosh Roy, Swagatam Das, Dhaval Jain, and Ajith Abraham, An Improved Harmony Search Algorithm with Differential Mutation Operator, Fundamenta Informaticae, IOS Press, vol. 95, no. 2009, pp. 1–26, 2009, DOI 10.3233/FI-2009-181.