International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimizing Cryptocurrency Price Prediction: A Hybrid Approach with Resilient Stochastic Clustering and Gravitational Search Algorithm

[1]R. Ramesh, [2]M. Jeya Karthic

Submitted: 20/12/2023 **Revised**: 31/01/2024 **Accepted**: 10/02/2024

Abstract: Cryptocurrency markets have become increasingly complex, making accurate price prediction a challenging task. This article proposes a Hybrid Oppositional Sparrow Search of Gravitational Search Algorithm (HOSS-GSA) which separated iterative chaotic routing to address problems of its probability of falling optimal solutions. The proposed hybrid framework aims to harness the strengths of each component to improve prediction accuracy and capture the dynamics of cryptocurrency historical price data. Resilient Stochastic Clustering effectively identifies relevant features and reduces dimensionality, enhancing the efficiency of subsequent prediction steps. Furthermore, it helps in identifying clusters of similar data patterns within the cryptocurrency historical prices dataset. HOSS-GSA aims to optimize model parameters and improve the overall performance of the prediction model. The experiments were conducted by the common evaluation operations to validate the functionality of grouping for high-dimensional multiview information of Cryptocurrency Historical Prices Dataset, as well as the Wilcoxon rank-sum assessment model was used to measure variable influence for the technique, outperforms traditional prediction methods, achieving higher prediction accuracy and robustness. This approach provides a valuable tool for cryptocurrency traders, investors, and analysts seeking to make informed decisions in a rapidly evolving market.

Keywords*: Oppositional Sparrow Search; Gravitational Search Algorithm; Folded Iterative Chaotic Mapping; Swarm Intelligence; Resilient Stochastic Clustering; Cryptocurrency Historical Prices Dataset.*

1. Introduction

Artificial Intelligence (AI), mobile services, media platforms, and the Internet of Things (IoT) are currently generating enormous amounts of information, resulting in the rapid advancement of big data applications. Big data, which would be characterized by volume, diversity, speed, truthfulness, and value, should be complicated and voluminous that standard Cryptocurrency Historical Prices data technological solutions are unprepared to cope with it. One of the most difficult in big data environments was analyzing high-dimensional multiscreen information of diverse complicated and huge-scale uses [1-2]. Cryptocurrency markets have experienced unprecedented growth and volatility in recent years, attracting significant attention from investors, traders, and researchers. Accurate price prediction in these markets is crucial for making informed investment decisions and managing risks effectively. However, the complex and dynamic nature of cryptocurrency price movements presents a formidable challenge for traditional prediction methods.

Deep Convolutional Neural Networks (DCNN) and varied architectures, collected from various sources, are frequently used to describe high-dimensional multi-screen Cryptocurrency Historical Prices data. Grouping multiview information would be an NP-hard issue in particular, which has prompted a slew of academics to offer a variety of clustering techniques to a variety of real-world situations. The group structure was kept for unsupervised feature selection of multiview tasks, and an interchanging technique was given to realize the architecture [3]. For multi-view aggregating, a novel multi-view alignment dissemination technique has been developed, which would be particularly suited for grouping more than two aspects. Grouping high-dimensional multi-view data differs significantly from standard clustering methods, which treat all images as a flat collection of variables and do not consider multiview to be an important component in segmentation [4]. Furthermore, because of its complex structure, multiple sources, and large volumes, a grouping of huge-dimensional multiscreen information was more difficult and crucial to the processing product's effectiveness [5]. Apache Spark, as a quick and generic motor of massive information analysis, could provide lightning-quick cloud infrastructure for aggregating high-dimensional multi-view data to implement a variety of huge information implementations, picture separation, web page categorization, clinical diagnostics, etc. (PSO) was commonly utilized to increase the efficiency of responses to a variety of research challenges [6-7]. To solve energy-efficient resource provisioning challenges, researchers created hybrid quantum particle swarm optimization.

[1]*Assistant Professor, Department of Computer and Information Science, Annamalai University, Annamalai nagar.*
rameshau04@gmail.com
[2]*Assistant Professor, Department of Computer and Information Science, Annamalai University, Annamalai nagar.*
Jeya_karthic@yahoo.com

In response to these challenges, this paper introduces a novel approach that leverages the power of three cutting-edge techniques: Resilient Stochastic Clustering (RSC), Hybrid Oppositional Sparrow Search (HOSS), and the Gravitational Search Algorithm (GSA). This innovative fusion of methods is designed to enhance the accuracy and robustness of cryptocurrency price prediction, addressing the unique characteristics of the cryptocurrency historical prices dataset.

Cryptocurrency markets are characterized by high volatility, non-linearity, and rapid fluctuations in price. Factors such as market sentiment, regulatory changes, and technological developments can have a profound impact on prices. Traditional financial models often struggle to capture the complexity and speed of these dynamics, necessitating the development of more sophisticated prediction techniques. RSC is employed as the first component of our framework to preprocess the cryptocurrency historical prices dataset. This technique is instrumental in identifying relevant features and clusters within the data, reducing dimensionality, and enhancing the quality of input data for subsequent analysis. By grouping similar data patterns, RSC provides valuable insights into the underlying structure of the cryptocurrency market.

The second component, HOSS, represents a hybridization of two powerful optimization algorithms: Oppositional-based Learning (OBL) and Sparrow Search Optimization (SSO). HOSS is utilized to optimize model parameters, offering a balance between exploration and exploitation. This dynamic optimization process aims to fine-tune the prediction model for optimal performance, adapting to the rapidly changing cryptocurrency market conditions. GSA is the third component of our framework, applied to further refine the model parameters. Inspired by gravitational forces in celestial bodies, GSA optimizes the prediction model, ensuring it converges toward the best possible solution. This unique optimization algorithm enhances the accuracy of cryptocurrency price predictions by effectively adapting to market fluctuations.

The rest of the paper follows: section 2 brief discussions about related works, section 3 gives proposed model of the paper. Results and discussion for the cryptocurrency historical prices dataset is provided in section 4. Finally paper concluded in section 5.

2. Related Works

A limited movement PSO was developed to maximize support vector evaluation adjustable parameters for nonlinear regression analysis and prediction [8]. To enhance the power collection from array, an approach to actual displacement based on PSO is also proposed [9]. For many optimization applications, the above PSO methods have outside effectiveness. For several PSO algorithms, huge data presently poses both an advantage and a difficulty [10]. New PSO variations that can operate quickly on Spark are urgently needed to maximize the grouping of huge-dimensional multiscreen data of massive information applications [11].

An exception is an occurrence that differs so significantly that it was produced by a distinct process [12]. Noise is defined as anything that distorts the connection between an instance's characteristics and its category, and reports caused by non-systematic errors [13]. Outliers have been shown to considerably skew grouping findings, particularly when the obvious implication is that every data point must belong to a group [14]. Segmentation and feature extraction were described as efficiency problems requiring previous data on the number of extremists. A method of grouping and recognizing outliers at the same moment is useful for k-means aggregation [15]. An incremental approach for identifying and removing outliers based on visualization and hierarchical agglomerative approaches.

Resilience assessments aim to estimate the price of a network intrusion in terms of the level of problems created by the assault. The size of the essential assault set [16] is a popular expense metric. An assault set is a group of nodes whose removal causes a network to be disrupted, causing it to be divided into unconnected elements. A node-cut, a cut group, or dividers are terms used to describe an assault set [17]. It's also necessary to calculate the amount of destruction caused by the elimination of an assault set. The number of ensuing elements or the size of the largest surviving linked element were two common techniques to quantify disruption. All measures, it should be noted, yield significant assault groups that are expected to include inter-cluster boundary nodes, bridge nodes, and bottlenecks. Candidate groups are built using the elements generated by deleting the crucial assault set [18].

Massive data analytics should be a prominent study issue, as well as it shows an important role in numerous massive data implementations, thanks to the immediate growth of network computing, framework as well as providing Segmentation is a key data mining method in Big Data Analytics (BDA) [19]. To deal with the issues of big data, this review introduces the direction and development of clustering techniques. To highlight the implications of big data, some clustering methods were evaluated to conceptual and empirical approaches. For big data, an existing algorithm is employed as a method to improve mapping between chart grouping and data grouping [20].

Image segmentation and classifier algorithms have received much attention, but feature selection has received less research. Swarm Intelligence algorithms are examined in this paper in relation to their applications. To classify the data, this paper [21] makes use of the K-Nearest Neighbor algorithm. Swarm intelligence algorithms are benchmarked by comparing the results of the two algorithms that have been discussed.

A growing body of research has been conducted on how to predict the performance of cryptocurrencies because they are good investments. Conventional methods are difficult to use to capture financial time series relating to cryptocurrency. The optimal parameters for Artificial Neural Networks (ANNs) are a tedious task, even though they are the better alternative. In this paper [22], evaluated the efficacy of the hybrid ANN using Rao algorithm for predicting six popular cryptocurrencies by combining ANNs and Rao algorithms.

Investors, academicians, and researchers have been interested in trading on the stock market for a long time. Stock market data is intrinsically non-linear, which makes it difficult to perform linear analysis of the data [23]. There are many learning algorithms developed to study market behavior and enhance prediction accuracy; they have been optimized using swarm and evolutionary computations, and their global optimization capability with continuous data is exploited.

Blockchain has become a major research area in financial studies as a result of the growing interest that cryptocurrency has captured among financial scholars. Among the many types of data and an abundance of resources that are available in the cryptocurrency research community, the use of machine learning algorithms has become possible [24]. As of yet, no comprehensive review has been conducted on cryptocurrencies using machine learning that are able to represent the underlying algorithms.

Due to the boom in bitcoin and Blockchain technology use, investors are concerned about returns on investments and risks associated with financial products. This means that it is important to anticipate, in advance, what the probability of the return of a bitcoin is going to be. The method presented in this research article [25] is one of the most effective methods of predicting the returns on Blockchain financial products, which has been found to be one of the most effective methods available. A hybridization process is employed in order to increase the diversity of the population of Grey Wolf Optimization as a way of eliminating the problem of local optima and increasing the diversity of the population.

3. Proposed Work

This research proposes a novel intelligent Weighting HOSS-GSA technique to deal with the grouping of huge-dimensional multiscreen Historical Prices of Cryptocurrency information of diverse massive data implementations. An amount of connectivity between clusters is supplied in the clustering model in the first phase to increase group divergences. The group centers, ratings of perspectives, and values of characteristics were translated into a component notation in the second phase. Then, to achieve improved initial group centers, ratings of perspectives, and values for characteristics, the Chaotic Particle Swarm Optimization (CPSO) procedure is implemented. The next phase proposes a specific modification to increase CPSO's application programmers. HOSS-GSA correlates Linear Regression (LR) and Support Vector Machine (SVM) on three different computational functions of apache spark and solitary clusters to determine effectiveness as shown in fig 1. Our proposed model combines three advanced techniques to create a robust and accurate cryptocurrency price prediction system.

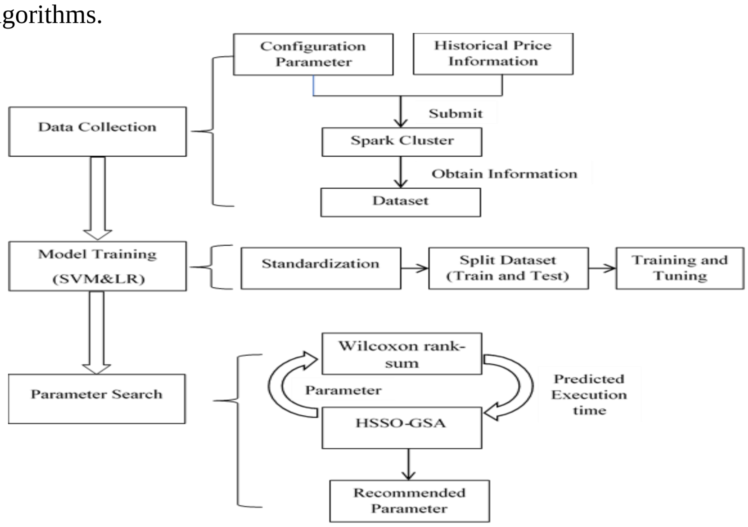

Fig 1: Overall Architecture of Proposed Model

3.1 Data Collection and Pre-processing:

Cryptocurrency historical price dataset is reputable sources which are collected from "https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory".Each currency is represented by a separate CSV file. Each day since April 28, 2013, it has been providing price history. Statistical information about some of the top cryptocurrency prices by market capitalization can be found in this dataset. A comprehensive dataset of historical cryptocurrency prices are presented including data points such as opening price, closing price, high and low prices, trading volume, and time stamps.

Table 1: Information about Historical Price

$Date: \ date\ of\ observation$
$Open: \ Opening\ price\ on\ the\ given\ day$
$High: \ Highest\ price\ on\ the\ given\ day$
$Low: \ Lowest\ price\ on\ the\ given\ day$
$Close: \ Closing\ price\ on\ the\ given\ day$
$Volume: \ Volume\ of\ transactions\ on\ the\ given\ day$
$Market\ Cap: \ Market\ capitalization\ in\ USD$

Historical price data in the cryptocurrency market dataset refers to the recorded values of a cryptocurrency's price over a specific period of time in the past. This data is essential for various cryptocurrency-related analyses, trading strategies, investment decisions, and research. The dataset undergoes thorough cleaning to handle missing values, outliers, and inconsistencies. High-quality data is essential for reliable predictions. Identify and handle missing data points, which may arise due to gaps in the historical record. This can interpolate missing values, fill them with zeros, or use other appropriate methods. Detect outliers or anomalies in the data. Outliers can significantly impact model performance. Decide whether to remove, transform, or impute outliers based on their nature and impact on analysis.

3.2 HOSS-GSA

This hybridization aims to enhance the efficiency and effectiveness of optimization processes in various applications, including machine learning, engineering, and finance. The following are the specific application stages:

Step 1: Create the citizens and their factors, including the citizen's size N, the percentage of innovators PD, the composition of officers SD, the impartial substring dimension D, the top and bottom confines of the original amount established to lb and ub, the largest number to repetitions T, as well as warning criterion ST.

Step 2: Create N*D-dimensional matrices Z_i, but also corresponding map to the character of person targeted area. This guarantees that the sparrow numbers remain diverse.

Step 3: Determine each sparrow's fitness f_i, choose the optimal fintness f_i but also associated position x_b, and the present lowest fw and associated location xw.

Step 4. Using the predetermined percentage PD, choose pNum sparrows to exceptional adaptation as explorers and the remainder as entrants at random. Modify to inventors' positions using of equation (1)

$$I_{x,y}^{t+1} = \begin{cases} I_{x,y}^{t}.EXP\left(\frac{x}{\alpha.Maximum}\right), r_2 < ST \\ X_{x,y}^{t} + P.L, \qquad r_2 \geq ST \end{cases} \quad (1)$$

Step 5: Enhance the latest global lowest score using the generalized opposition-based learning method.

Step 6. Using the ratio SD, select sNum guards at random from the population and execute the horizontal crossover and activity.

$$i(t+1) = r.i(t).\left(1 - i(t)\right), r \in N, i(0) \in [0,1] \quad (2)$$

$$d(pBest, gBest) = \frac{1}{N.S}\sum_{x=1}^{N.S}\sqrt{\sum_{y=1}^{dim}\left(pBest_{xy}, gBest_y\right)^2} \leq \in.d \quad (3)$$

Step 7: Conduct a vertical crossing procedure using the equation below, assess fitness levels, and save the better ones.

$$sd\left(pBest_j\right) = \sqrt{\frac{1}{N.S}\sum_{x=1}^{N.S}\left(pBest_{xj} - \frac{1}{N.S}\left(pBest_{1y} + pBest_{2y} + \cdots + pBest_{Nsy}\right)\right)^2} \leq \quad (4)$$

Step 8. Throughout the foraging operation, modify the optimal location xb, the best health value fb, the terrible location xw, and the horrible standard evaluation fw of the overall population depending on the present status of the sparrow community.

Step9: Verify that the iteration was completed. When the method reaches the maximum number of incarnations or the solution data reaches the set point, the loop ends, and the optimal control outcome is printed. However, it continues to Step 2 begins the next incarnation function, and the latest incarnation amount t meets the condition t = t + 1.

Step 10: Print to HOSS-GSA statistics.

Grouping was modeled optimization of the following objective function for separating X into C groups with values of views and characteristics.

$$itness(U, Z, WV, WF) = \frac{\sum_{l=1}^{C}\sum_{x=1}^{N}\sum_{t=1}^{T}\sum_{y\in Viewt} u_x wv_t wf_y (i_{x,y}-z_{ky})^2}{\sum_{k=1}^{C}\sum_{t=1} T \sum_{y\in Viewt} wv_t wf_y (z_{k,y}-o_y)^2} \quad (5)$$

$$.ST= \begin{cases} \sum_{t=1}^{T} wv_t = 1, 0 \le wvp_t \le 1 \\ \sum_{k=1}^{C} U_{x,k} = 1, 1 \le x \le N, U_{x,k} \in [0,1] \\ \sum_{y\in Viewt} wf_y = 1, 0 \le wf_y \le 1, 0 \le t \le T \\ o_y = \sum_{k=1}^{C} {z_{k,y}}/{C} \end{cases}$$

(6)

3.3 Prediction classifiers

Prediction classifiers, often referred to as machine learning classifiers or predictive models, are algorithms used to classify or categorize data into different classes or categories based on input features. LR and SVM classifiers are used as prediction classifier for analysing historical price data. The predictor variable used to train and validation the database was success or failure. 75 percent of the database was utilized to build the machine, while the remaining 25% was being used to evaluate the model.

- Allocate a value of 0 to failure and 1 for success to the target variable p.
- The constant value was p0.
- The exponential base integer is b.

$$L_0 = b^{Po+P\sum_{x=1}^{i} fx} \quad (7)$$

$$\gamma = \frac{L_0}{\left(b^{Po+P\sum_{x=1}^{i} fx}\right)+1} \quad (8)$$

$$= \frac{1}{1+b^{-\left(b^{Po+P\sum_{x=1}^{i} fx}\right)}} \quad (9)$$

Between [0,1] is the likelihood probability. If this number was larger than 0.5, the company's pre-launch forecast was regarded as an achievement, while numbers less than 0.5 are judged a failure in this study. SVM stands for supervised machine learning, which would be a technique for acquiring new skills and cryptocurrency data from a group of statistics. The training process for SVMs involves finding the optimal hyperplane that best separates the data points belonging to different classes while maximizing the margin. The mathematical equations involved in the training process are given below,

Data Representation:

Let's assume you have a training dataset with n samples and m features:

Input Features: $X_1, X_2, X_3, \ldots, X_m$

Target Labels: y (binary: -1 or 1 for a binary classification problem)

Objective Function:

The SVM aims to maximize the margin while minimizing classification errors. The margin is defined as the distance between the decision boundary (hyperplane) and the support vectors. The objective function to optimize is:

$$Maximize: 2 / ||w|| \quad (10)$$

Subject to: $y_i(w \cdot x_i + b) \ge 1$ for all i, where w is the weight vector, x_i is the data point, b is the bias term, and y_i is the class label for data point x_i.

Soft Margin SVM:

In practice, it's often not possible to find a perfect separation, so a "soft margin" approach is used. This introduces a parameter C that controls the trade-off between maximizing the margin and minimizing the classification errors. The objective function becomes:

$$Minimize: 0.5 * \left||w|\right|^2 + C * \Sigma\xi_i \quad (11)$$

$$Subject\ to: y_i(w \cdot x_i + b) \ge 1 - \xi_i\ and\ \xi_i \ge 0 \quad (12)$$

for all i, where ξ_i are slack variables.

Optimization Problem:

The SVM training problem becomes an optimization problem to find w and b that minimize the objective function while satisfying the constraints. This often involves solving a convex quadratic programming problem.

4 Results and Discussions

The provided dataset exemplifies a conventional format for a historical price dataset, particularly within the realm of financial markets and cryptocurrencies. Every every data point contains valuable information on the trading activity of a certain asset, such as a stock or a cryptocurrency, on a given day. Each data field is comprised of:

Date: The date serves as a reference point for the observation, denoting the particular day on which the remaining data points were documented. The inclusion of the date is of utmost importance, as it serves to effectively arrange the data in a chronological manner.

Open: The opening price refers to the initial value of an asset at the commencement of the trading day. The reference point facilitates comprehension of the fluctuations in the price of the item over the course of the day.

High: The term "high" refers to the maximum trading price seen for a certain item during a single trading day. The data indicates the highest point attained by the asset during the given day.

Low: The term "low" refers to the minimum trading price seen for an asset throughout the course of a trading day. This denotes the lowest point attained by the asset throughout the given day.

Close: The closing price refers to the value of the asset at the conclusion of the trading day. The assessment of an asset's performance during a trading session has significant importance.

Volume: The volume of transactions refers to the aggregate quantity of shares or units of the asset that have been exchanged during a certain trading day. A high volume of trade often signifies heightened trading activity, which may have a substantial influence on price fluctuations.

Market Cap: Market capitalization, sometimes referred to as market cap, represents the aggregate value of all outstanding units of a certain asset that are currently in circulation. This metric is typically denominated in United States dollars (USD). The calculation involves the multiplication of the prevailing price by the aggregate quantity of units in circulation. The market capitalization is a fundamental indicator used to evaluate the magnitude and comparative worth of an item within a given market.

This dataset is commonly used for various financial and investment analyses, including technical analysis, fundamental analysis, and risk assessment. It provides a historical record of how the asset's price and trading volume have changed over time, helping analysts and investors make informed decisions based on past market behavior.

4.1 Experimental setup

The Apache Spark 2.2.1 architecture was used to create the proposed methodology. The program was created using Python's Spark software and PySpark 2.1.2, which would be the Spark Python API. The Web Application Gateway Connection should be used with an Ubuntu internet server running Apache. On several Spark cluster arrangements, Amazon Network Services was utilized to operate various components of the program application on big servers with two Intel Xeon E5-2699V4 2.2 G Hz computers (VCPUs) of 4 cores and 16 GB of RAM. Software components could be set up and executed on multiple servers based on sustainability needs. HOSS-GSA could handle big data applications and achieve an outside achievement using the Apache Spark machine learning method and RDD, as demonstrated in Figure 2.

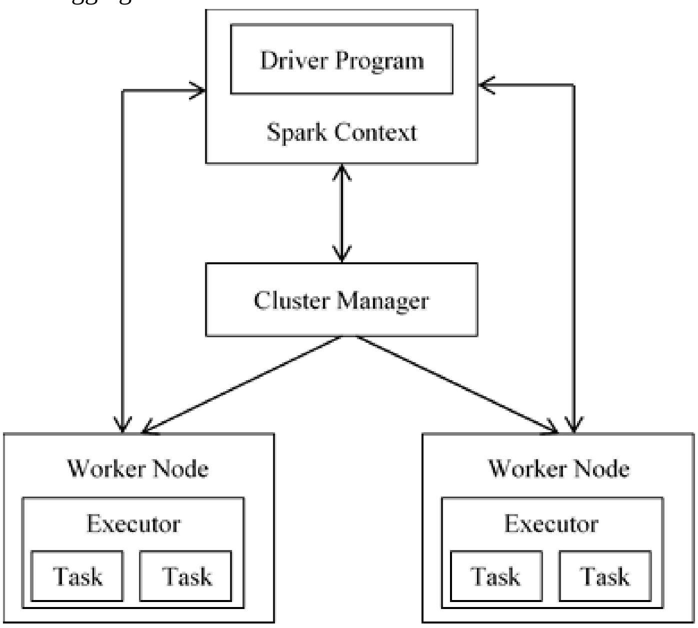

Fig 2: HOSS-GSA on Apache Spark

4.2 Performance Evaluation

Numerous case studies were designed to validate our prediction system. The probability was calculated using SVM and LR algorithms. The system's performance is assessed using the most essential components. The network design parameters are the model performance assessment. Another key challenge for huge data processing systems was network reaction time. To find important attributes and remove duplicates from the Cryptocurrency Historical Prices dataset, they propose using component Historical Prices data gain and the HOSS-GSA technique. Figure 3 depicts key characteristics essential to the long-term viability of historical prices of cryptocurrency distribution plots for all the features using Kernel Density Estimation, which showed that the data seemed to be highly skewed. The graph represents the designer's important characteristics versus the quantity of cryptocurrency data studied. External memory was also mentioned as a key component. This could help to improve the Cryptocurrency Historical Prices data greater product quality and business stability.

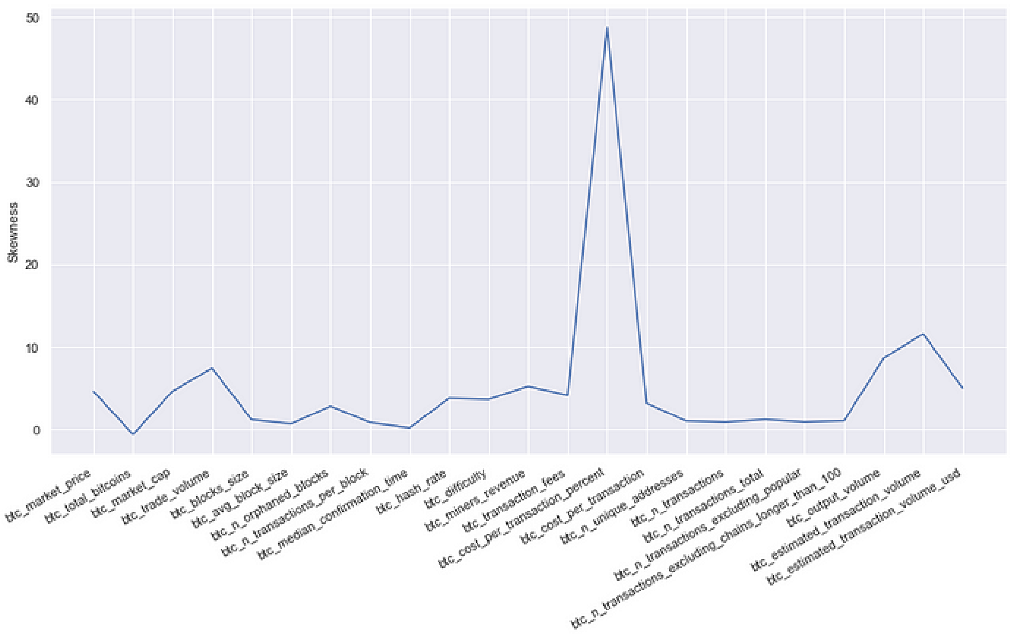

Fig 3: Important Feature of Skewness of Dataset

Figure 4 compares the computing time required by the proposed models of different Cryptocurrency Historical Prices database sizes for the production time consumed by state-of-the-art approaches. The HOSS-GSA procedure employs less time to perform the software to Gini-index and implicit semantics evaluation techniques. As a result, the proposed technique would be both quick and adaptable. With big databases, it gives high-speed processor performance. This demonstrates the feasibility of HOSS-GSA in big data analytics, whereas the waiting period for Gini-index and Latent Semantic Analysis (LSA-based approaches) was longer for Cryptocurrency Prices data. The time spent to forecast using the LSA-based method, Gini-index prototype, and HOSS-GSA framework with a 9 GB database was 156 s, 495 s, and 342 s, correspondingly. With an 18 GB database, the LSA model, Gini-index prototype, and HOSS-GSA framework took910 s, 256 s, and, 740 s to forecast, correspondingly. The time required for Gini-index and LSA algorithms of an 18 GB database was twice of a 9 GB database. However, the time consumed by the HOSS-GSA method for an 18 GB database was 1.6 times of a 9 GB database, and it should be 3 times faster than the Gini-index technique. In terms of application implementation and effectiveness, the HOSS-GSA model outperforms other state-of-the-art approaches.

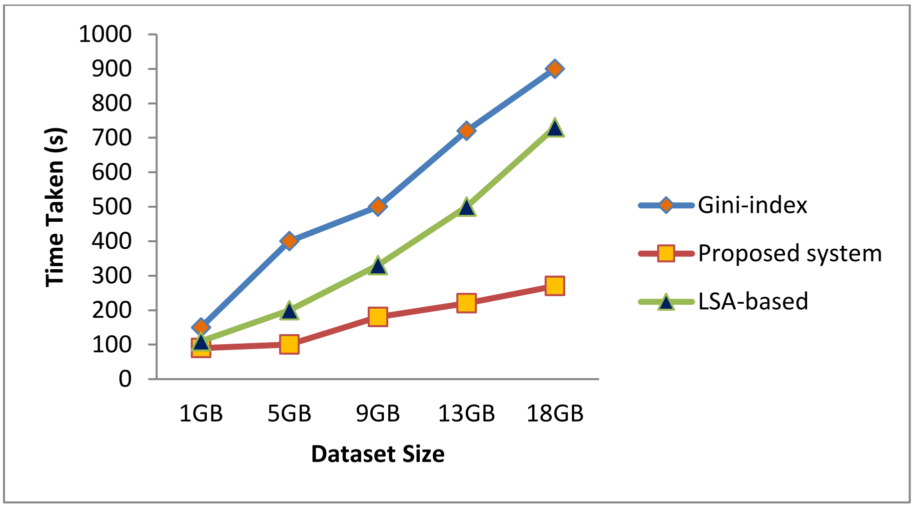

Fig 4: Comparison of size and Production Time

Figure 4 compares the proposed Gini-index and LSA approaches to Linear Regression and Support Vector Machine classification over 24 months in terms of precision, memory, and efficiency. The shows in Table 3 were great demonstrated utilizing HOSS-GSA with SVM classification, which has a 95.4 percent predictive performance. In Recall (R), Accuracy (A) and Precision (P) evaluations, the HOSS-GSA exhausted LSA-based and Gini-index techniques. True positive (TP), false-negative (FN), true negative (TN), and false-positive (FP), were identified using the proposed approach. Eqs. (13), (14), and (15) are used to calculate P, R and A respectively.

$$A = \frac{\text{True Positive+True Negative}}{\text{True Positive+True Negative+False Positive+False Negative}} \quad (13)$$

$$P = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (14)$$

$$R = \frac{\text{True Positive}}{\text{True Positive+False Negative}} \quad (15)$$

Table 1 Proposed model compares to state-of-the-art methodologies in terms of capability.

Classifier Method	**SVM**		
	P%	**R%**	**A%**
HSSO-GSA	0.951	0.93	95.6
LSA-based	0.896	0.80	87.6
Gini-index	0.67	0.571	83.3
Classifier Method	**LR**		
	P%	**R%**	**A%**
HSSO-GSA	0.917	0.852	94.2
LSA-based	0.839	0.764	84
Gini-index	0.63	0.53	81.2

When contrasted to LSA-based and Gini-index techniques, the prediction differences using HOSS-GSA to SVM classifier and LR classifier are fewer. As a result, for component forecasting, HOSS-GSA surpasses the two approaches. Several aspects of HOSS-GSA may occur in multiple ways, summative assessment would not take into account repeated evaluations. Characteristics are retrieved depending on the polarity of the evaluation in the Gini-index, R and P are lower for large datasets. The findings show that in big data analysis, the HOSS-GSA method surpasses the other 2 techniques. In characteristic forecasting, the Gini-index method fails miserably.

4.2 Wilcoxon Rank Sum Test

The summative assessment of enhanced smart optimization techniques, based just on a median, average, and optimal quantities, was insufficiently persuasive. The accuracy of the correlation test findings, which is one of the major components, also indicates if the technique has been greatly improved or not. The Wilcoxon statistical technique used a significance value of 5% to determine whether the findings of the enhanced HOSS-GSA in this research are substantially contrary to the findings of other techniques. The following is a brief description of the assessment concept: When the P-value was much less than 0.05, the difference between the two methods was considered important. P 0.05, the performance of the two methods was comparable, and the variation is discernible. The complete significance level > 0.05 is given as N/A in this section. The P-value computed in the Wilcoxon rank-sum assessment of HOSS-GSA and other techniques among the 14 benchmark datasets is shown in Table 2. The primary elements are explained by P 0.05, according to the results. The HOSS-GSA method outperforms the SSA approach, and the difference is statistically significant, indicating that the upgraded method has good convergence precision.

Table 2: p-value and Wilcoxon

Function	LR	SVM	HSSO-GSA
F1	3.02E-11	3.01E-11	0.00654
F2	1.21E-10	1.19E-10	7.48E-9
F3	3.03E-11	3.02E-11	8.99E-10

The results in the table represent the performance of each function (F) when assessed with these different algorithms. The term "function" refers to different mathematical functions or algorithms that are being evaluated or tested.

5 Conclusions

This article suggests the construction of a HOSS-GSA method that incorporates the cryptocurrency data increase and Shares Memory-based Resilient cryptocurrency Database Filter methods with LR and SVM forecast predictors to identify relevant characteristics and remove unnecessary cryptocurrency data. The HOSS-GSA technique exceeds the other techniques in classification performances, precision, but also memory. When contrasted to state-of-the-art methodologies, the proposed technique improves prediction performance to 10% by identifying significant features and removing repetition from the cryptocurrency data. The LSA-based technique's predictive performance is reduced by large feature complexity, where the number of selected characteristics plays an important part in forecast modeling. The loading separation algorithm in HOSS-GSA would be used to identify the groups of items using varying values of views and features. Then CPSO calculates the first group's centers, view weights, and characteristic weights, which are improved to an accurate disturbance of the pBe.st and gBe.st. Results reveal that the proposed HOSS-GSA approach was adaptable and performs well with large datasets, and takes less time to analyze the program than state-of-the-art methodologies. A technique was predicted to have an influence on a grouping of great-dimensional multiview cryptocurrency data in a variety of high data applications. In future, develop a framework for online learning that continuously updates the model as new data arrives with Commercial Market Predictions Techniques using PMI scoring model.

References

[1] Gai, J., Zhong, K., Du, X., Yan, K., & Shen, J. (2021). Detection of gear fault severity based on parameter-optimized deep belief network using a sparrow search algorithm. *Measurement*, *185*, 110079.

[2] Jianhua, L., & Zhiheng, W. (2021). A hybrid sparrow search algorithm based on constructing similarity. *IEEE Access*, *9*, 117581-117595.

[3] Sharma, A., & Nanda, S. J. (2022). A multi-objective chimp optimization algorithm for seismicity de-clustering. *Applied Soft Computing*, 108742.

[4] Sharma, A., & Nanda, S. J. (2022). A multi-objective chimp optimization algorithm for seismicity de-clustering. *Applied Soft Computing*, 108742.

[5] Meng, Z., Zhang, Y., Zhu, B., Pan, Z., Cui, L., Li, J., & Fan, F. (2022). Research on rolling bearing fault diagnosis method based on ARMA and optimized MOMEDA. *Measurement*, *189*, 110465.

[6] Jiang, F., Han, X., Zhang, W., & Chen, G. (2021). Atmospheric PM2. 5 Prediction Using DeepAR Optimized by Sparrow Search Algorithm with Opposition-Based and Fitness-Based Learning. *Atmosphere*, *12*(7), 894.

[7] Wu, H., Zhang, A., Han, Y., Nan, J., & Li, K. (2022). Fast stochastic configuration network based on an improved sparrow search algorithm for fire

flame recognition. *Knowledge-Based Systems*, 108626.

[8] Yang, L., Li, Z., Wang, D., Miao, H., & Wang, Z. (2021). Software defects prediction based on hybrid particle swarm optimization and sparrow search algorithm. *IEEE Access*, *9*, 60865-60879.

[9] Sun, C. J., & Gao, F. (2022). Remote Sensing Image Recognition Based on LOG-T-SSA-LSSVM and AE-ELM Network. *Computational Intelligence and Neuroscience*, *2022*.

[10] Chen, H., Ma, X., & Huang, S. (2021, August). A Feature Selection Method for Intrusion Detection Based on Parallel Sparrow Search Algorithm. In *2021 16th International Conference on Computer Science & Education (ICCSE)* (pp. 685-690). IEEE.

[11] Chang-Jian, S., & Gao, F. (2022). Remote Sensing Image Recognition Based on LOG-T-SSA-LSSVM and AE-ELM Network. *Computational Intelligence and Neuroscience: CIN*, *2022*.

[12] Zhu, D., Xie, L., & Zhou, C. (2022). K-Means Segmentation of Underwater Image Based on Improved Manta-Ray Algorithm. *Computational Intelligence and Neuroscience*, *2022*.

[13] Pacifico, L. D., & Ludermir, T. B. (2021). An evaluation of k-means as a local search operator in hybrid memetic group search optimization for data clustering. *Natural Computing*, *20*(3), 611-636.

[14] Kumar, T. S., Arun, C., & Ezhumalai, P. (2022). An approach for brain tumor detection using optimal feature selection and an optimized deep belief network. *Biomedical Signal Processing and Control*, *73*, 103440.

[15] Abualigah, L. (2021). Group search optimizer: a nature-inspired meta-heuristic optimization algorithm with its results, variants, and applications. *Neural Computing and Applications*, *33*(7), 2949-2972.

[16] Nguyen, A., Roth, N., Ghassemi, N. H., Hannink, J., Seel, T., Klucken, J., ... & Eskofier, B. M. (2019). Development and clinical validation of inertial sensor-based gait-clustering methods in Parkinson's disease. *Journal of neuroengineering and rehabilitation*, *16*(1), 1-14.

[17] Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems Science & Control Engineering*, *8*(1), 22-34.

[18] Yuan, J., Zhao, Z., Liu, Y., He, B., Wang, L., Xie, B., & Gao, Y. (2021). DMPPT control of photovoltaic microgrid based on the improved sparrow search algorithm. *IEEE Access*, *9*, 16623-16629.

[19] Ma, B., Lu, P., Zhang, L., Liu, Y., Zhou, Q., Chen, Y., ... & Hu, Y. (2021). Enhanced Sparrow Search Algorithm With Mutation Strategy for Global Optimization. *IEEE Access*, *9*, 159218-159261.

[20] Zhang, C., & Ding, S. (2021). A stochastic configuration network based on a chaotic sparrow search algorithm. *Knowledge-Based Systems*, *220*, 106924.

[21] Majid, M. A., Abidin, A. F. Z., Anuar, N. D. K., Kadiran, K. A., Karis, M. S., Yusoff, Z. M., ... & Rizman, Z. I. (2018). A comparative study on the application of binary particle swarm optimization and binary gravitational search algorithm in feature selection for automatic classification of brain tumor MRI. Journal of Fundamental and Applied Sciences, 10(2S), 486-498.

[22] Nayak, S. K., Nayak, S. C., & Das, S. (2021). Modeling and forecasting cryptocurrency closing prices with rao algorithm-based artificial neural networks: A machine learning approach. FinTech, 1(1), 47-62.

[23] Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization. Archives of Computational Methods in Engineering, 28, 2133-2164.

[24] Ren, Y. S., Ma, C. Q., Kong, X. L., Baltas, K., & Zureigat, Q. (2022). Past, present, and future of the application of machine learning in cryptocurrency research. Research in International Business and Finance, 63, 101799.

[25] Sivaram, M., Lydia, E. L., Pustokhina, I. V., Pustokhin, D. A., Elhoseny, M., Joshi, G. P., & Shankar, K. (2020). An optimal least square support vector machine based earnings prediction of blockchain financial products. IEEE Access, 8, 120321-120330.