Information fusion-based genetic algorithm with long short-term memory for stock price and trend prediction

Ankit Thakkar, Kinjal Chaudhari *

Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382 481, Gujarat, India

ABSTRACT

Information fusion is one of the critical aspects in diverse fields of applications; while the collected data may provide certain perspectives, a fusion of such data can be a useful way of exploring, expanding, enhancing, and extracting meaningful information for a better organization of the targeted domain. A nature-inspired evolutionary approach, namely, genetic algorithm (GA) is adopted for a variety of applications including stock market prediction. The complex, highly fluctuating financial market-related problems require optimized models for reliable forecasting. Also, it can be observed that stock market etiquettes are generally non-linear in nature and therefore, a broader understanding and analysis of such market behaviors necessitate the collection and fusion of relevant information based on different associated factors. In this article, we propose an information fusion-based GA approach with inter-intra crossover and adaptive mutation (ICAN) for stock price and trend prediction. Inspired by the genetic diversity and survival capability of various organisms, our proposed approach aims to optimize parameters of a long short-term memory prediction model, and selects a set of features; to address these problems of interest, we integrate inter-chromosome as well as conditional intra-chromosome crossover operations along with adaptive mutation to diversify the potential chromosome solutions. We illustrate the step-by-step procedure followed by GA with ICAN and evaluate its performance for one-day-ahead stock price and trend prediction. GA with ICAN-based optimization results in an average reduction of 43%, 27%, and 26% using mean squared error, mean absolute error, and mean absolute percentage error, respectively, as compared to the existing GA-based optimization approaches; further, an average improvement of 61% is encountered using R^2 score. We also compare our work with Ant Lion Optimization approach and demonstrate the significance of GA with ICAN-based optimization. We analyze statistical significance, as well as convergence functions, for GA with ICAN and discuss remarkable performance enhancement; we provide necessary concluding remarks with potential future research directions.

1. Introduction

Diversity can be viewed from numerous perspectives; the field of information fusion can be seen as the diversification of several aspects that are fused together [1]. With respect to its classification using one or more qualitative variables, diversity can be considered as "the position of a population along a continuum ranging from homogeneity to heterogeneity" [2,3]; one of the widely explored topologies has described diversity based on separation, variety, and disparity [4]. The significance of diversity among humans has also been explored over biological, neurological, behavioral, psychological, ideological, geographical, cultural, demographical, intellectual, social, educational, economical, spiritual, and many other aspects; individuals' adaptability to such differences has been an important field of research [5–7]. The persistence, progression, and problem-solving capabilities of various living organisms have encouraged the development of several computational approaches; genetic algorithm (GA) is one of such nature-inspired algorithms based on the concept of evolution [8]. It is a population-based stochastic algorithm with selection, crossover, and mutation as its main operators. A population of solutions undergoes a series of operations and successive generation(s) of solutions can be derived in order to find the fittest solution for the given problem [9]; here, the fitness can be calculated using the defined objective function. Thus, natural evolution can be simulated using GA and the same can be further applied to select specific values, optimize given variables, identify plausible solutions as well as models, and analyze a broad range of complex real-world problems [9–11].

* Corresponding author.
  E-mail addresses: ankit.thakkar@nirmauni.ac.in (A. Thakkar), 17ftphde21@nirmauni.ac.in (K. Chaudhari).

Financial markets are referred to be "a fascinating example of complexity in action" [12]. The highly fluctuating stock markets have been widely studied to understand the market volatility, trends and patterns, investors' attitudes, economic etiquettes, and several financial aspects; fusion has been one of the useful ways to extract essential market information [1]; also, a large number of individuals, as well as organizations, invest in such markets with a primary motive to have an increased return of the invested amount [13]. Due to a large number of influential events, it is important to analyze the market; though careful study and trading can benefit the traders with higher returns, loss of valuables may also be experienced with stock market investments [14]. The manual analysis of a considerably large amount of stock data can be difficult, and therefore, various analytical tools, as well as computational techniques, have been developed to derive useful market information and potential hidden patterns. The nature-inspired algorithms have been extensively adopted for a wide range of applications; such metaheuristic approaches have also been applied to various computational methods to enhance the prediction performance of financial markets [15]. GA is one of the important demonstrations of natural evolution; the ability to survive through various environments and critical conditions has been an inspiration to address a large number of computational problems. It can be observed that diversity is crucial for any organism, and similarly, the algorithmic representation of a given problem can be enhanced using diversification. With an inspiration to integrate genetic diversity within a computational approach through information fusion, this article aims to enhance the prediction model's ability to forecast stock price and stock trend.

The stock market predictions are complex real-world problems; the prediction performance can be largely dependent on market analysis, collected and derived information, a potential fusion of the available data, identification of a set of relevant and informative features, as well as selection of a prediction model along with tuning of its hyperparameters. While several metaheuristics have been proposed for one or more of such applications, in this article, we propose to apply information fusion to GA based on inter-intra crossover and adaptive mutation (ICAN) for stock price and trend prediction. We propose to address two of the largely considered aspects in prediction viz. parameter optimization [16] and feature selection [17]. We integrate these two problems within a single chromosome representation and apply GA with ICAN to optimize the solution; along with the selected set of features, the optimized parameters are given to a prediction model, namely, long short-term memory (LSTM), so as to predict one-day-ahead stock close price trend. Here, it must be noted that the length of a chromosome shall be determined by the target application specifications such as the number of parameters to be optimized and the number of features to be selected. In such scenarios, the processing time would be dependent on chromosome length as well as other GA parameters. Here, articles [16,17] are referred for comparison with the proposed approach because of their potential applications of stock market prediction using GA-based parameter optimization and feature selection; while the existing approaches focus on one of the two aspects, we combine these two problems along with the proposed GA with ICAN. Our aim is to demonstrate the significance of genetic diversity in GA-based computational approaches, and therefore, we set simulation parameters as given by the referred existing approaches [16,17] for an unbiased evaluation and comparison of our proposed approach. While we introduce information fusion-based diversity in GA, its extension to other related metaheuristic approaches can be a potential research direction [18,19].

1.1. Motivation

Genetic diversity is considered to be an essential factor for adaptability, extensibility, and survival of populations [20]. While zygosity refers to the extent to which the chromosome parts have the same gene sequences, heterozygosity introduces the difference in the same; genomic heterozygosity is an important aspect to maintain genetic differences over reproduction. In contrast to that, another critical concept is inbreeding which refers to the mating of individuals that are closely related to each other based on their genetics. These factors play crucial roles in determining the fitness of the corresponding population and hence, their survival over a longer duration [21]; such diversities are highly dependent on the mating system, evolutionary history of a species, population history over time, as well as the levels of environmental heterogeneity [21]. Subsequently, several studies show that diversity of parental environments can increase the phenotypic variation, however, productivity can also be affected [22]. Therefore, it is of critical importance to maintain diversity but it should not be exceeding too far to degrade the productivity [22].

In GA, several approaches have been proposed to maintain diversity such as nitching, crowding, restricted mating, elitist, injection, to name a few [23]. To introduce genetic diversity with the primary goal of increasing the problem-solving capability of GA, we take motivation from the concept of conditional diversification; the considered chromosome parts representing parameter optimization and feature selection generally undergo inter-chromosome crossover such that individual parts are operated under crossover, separately. We establish conditions with an aim to take out GA from being potentially stuck within local optima; for this purpose, an intra-chromosome crossover operation is additionally applied before the regular inter-chromosome crossover in case certain conditions are met. In other words, we define conditions and keep a check on the chromosomes' performance over subsequent generations; in case of meeting the defined condition, one of the two chromosomes undergoes intra-chromosome crossover, followed by an inter-chromosome crossover with the other chromosome for information fusion. This conditional intra-chromosome crossover is brought in for exploration purposes. Further, we propose to adaptively mutate the genes to maintain diversity in accordance with the network's fitness. This can be viewed as a method to evaluate change in performance of the best chromosome and determine a potential way to increase exploitation; as exceeding diversity may degrade productivity [22], we have designed adaptive mutation to maintain exploitation through generations. Thus, the primary motivation is to build an information fusion-based GA approach that can overcome the potential limitations of GA by introducing diversity through ICAN and stimulating exploration, as well as exploitation, through the proposed conditions. Results indicate that a significant improvement in prediction performance is obtained by GA with ICAN-based parameter optimization and feature selection approach in predicting stock price and trend. Further, the proposed approach is compared with a metaheuristic approach, namely, Ant Lion Optimization (ALO); the results also showcase higher prediction performance using GA with ICAN.

The major contributions of this article are as follows.

- We prepare each chromosome with two parts to address parameter optimization and feature selection tasks in parallel wherein we apply inter-chromosome crossover for individual parts and introduce an information fusion-based conditional intra-chromosome crossover operation for genetic diversity.

- We propose to update the mutation rate adaptively by evaluating the fitness values in the previous and current generations and evaluate the impact of genetic algorithm with inter-intra crossover and adaptive mutation (ICAN) for stock price and trend prediction.

The remaining article is organized as follows: the literature related to the proposed work is briefly reviewed in Section 2; a detailed procedure of our proposed approach along with an illustrative example is provided in Section 3; the simulation parameters and experimental results are analyzed and discussed in Section 4; remarks on GA with ICAN for stock price trend prediction are concluded in Section 5.

2. Related work

The financial markets consist of extensive operations and transactions over numerous assets for which market forecasting, portfolio optimization, risk analysis, etc. can be carried out. With a rise in computational intelligence techniques, such applications have gained major attention from researchers of various disciplines. While machine learning or deep learning models can be utilized for computational finance-related problems, it is of critical importance to identify an optimal way to build a model for performance enhancement. Integration of metaheuristics can be useful for feature selection, model optimization, forecasting tasks, to name a few. Inspired by genetic diversity, we propose an information fusion approach with GA for stock price trend prediction; to understand the potential applications of GA in the existing literature, this section briefly reviews recent related work based on GA and financial applications.

Artificial neural network (ANN) models are significantly used with a variety of datasets including stock market indices. A potential enhancement to an ANN model can be based on its weights and biases; authors in Ref. [24] integrated GA with ANN for the model's weight optimization, validated the model using COVID-19 stock data, and predicted next day stock indices using the hybrid model. On the other hand, GA regularization online extreme learning machine (ELM) was developed with a forgetting factor to estimate the effects of news during the COVID-19 pandemic on the crude oil futures volatility in Ref. [25]. Other applications of GA-based parameter tuning include variational mode decomposition (VMD) model's optimization for data sequence decomposition and LSTM-based prediction [26], adaptive neuro-fuzzy inference system (ANFIS) optimization for stock market index prediction [27], perceptron neural network optimization for determining effective features to detect fraud in financial statements [28], to name a few. In Ref. [29], GA was utilized to construct an optimal period and a simple average was conducted on the trading volume during such period to forecast trading volume using dynamic time warping (DTW). One of the stock market phenomenons is post-earnings-announcement drift (PEAD) which refers to the tendency of stock's cumulative abnormal return to drift after an earnings announcement; to study PEAD dynamics, authors proposed to optimize extreme gradient boosting (XGBoost) using GA in Ref. [30]. Another approach developed a spatio-temporal convolutional neural network model for learning complex correlations between time-series data of financial markets [31]; authors integrated GA with a constrained gene for identifying abnormal situations for outlier companies in the model fit.

Potential GA applications can be viewed in portfolio optimization; Google's Bidirectional Encoder Representations from Transformers (BERT) model was adopted for capturing market conditions through sentiment analysis using tweets wherein GA was applied for model optimization based on different objectives in Ref. [32] such that the proposed approach improved the performance of a portfolio. Another approach addressed the cardinality constraint portfolio optimization using hybridized nonlinear neural network and GA in Ref. [33] whereas a hybrid intelligent algorithm was developed based on mean–variance-skewness portfolio optimization using GA under an uncertain environment in Ref. [34]. Subsequently, GA chromosomes were encoded with stop-loss and take-profit (SLTP) points as well as group trading strategy portfolio (GTSP) along with strategy group weights with an aim to determine reliable GTSP and appropriate SLTP in Ref. [35]. The natural selection-based GA approach was also proposed for determining safe stocks, i.e., stocks indicating safe long-term investments, using the rank-based manner in Ref. [36]. Among several GA-based optimization approaches, LSTM model can be seen widely utilized for complex temporal stock market data due to its ability to deal with long-term dependencies; in Ref. [16], authors determined LSTM parameter values through GA-based optimization whereas in Ref. [17], authors selected a set of optimal features using GA and applied the features for stock prediction using LSTM model. Another such approach introduced adaptive GA for automatic optimization of network structure and hyperparameters in Ref. [37].

It can be viewed that the applicability of GA-optimized model parameters, selected features, as well as other forecasting methods, denote the usefulness of GA in financial applications. While direct implications of GA have been observed, it is important to evaluate its generalization ability in forecasting different types of data [24]; such data may have variable characteristics, different value ranges, and other dependencies that may require further generalization. While the experiments may be conducted on specific stock market indices, their expansibility to other markets must be evaluated to showcase adaptability and generalization ability [29]. Similarly, having vast differences in the prediction performances of a few indices or sectors may denote limitations in the features [30]; such aspects require further attention in order to prepare a set of features that can significantly determine useful properties of the considered data. Also, the consumption of available resources due to accessive computation is challenging [26]; it can be viewed from the criticality perspective to understand the necessity of time and space complexities associated with the approaches. For applications that require critical monitoring, the performance of an optimization approach requires to be evaluated based on time-specific metrics too. Subsequently, working with non-linear time-series markets can increase the challenges due to several many factors that can temporarily and/or permanently affect the markets [27].

Due to the potential of GA, it is of critical importance to analyze its internal operations that can further contribute to expanding its problem-solving capabilities. Among several existing approaches, limited attention is provided to introducing genetic diversity. To the best of our knowledge, this article can serve as a pioneering approach to conditionally diversify GA's evolution process for enhancing exploration as well as exploitation criteria. We apply our proposed information fusion-based GA with ICAN approach to one of the complex real-world applications, i.e., stock price and trend prediction; a step-by-step discussion on the proposed work along with an illustration and experimental evaluation are provided in the subsequent sections.

3. Methodology

Nature has inspired an enormous number of ways to address diverse optimization problems. One of such algorithms is GA based on the scalability and searchability in large search spaces [8]. In this article, we aim to integrate genetic diversity to enhance the computational approaches; the diversified solutions are further evaluated for stock trend prediction.
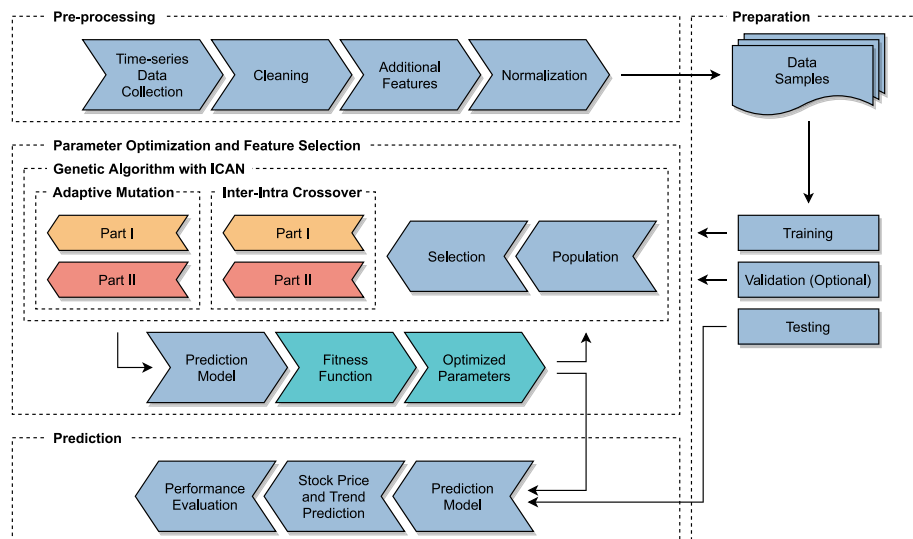
Fig. 1. Structural overview of the proposed approach.

3.1. Proposed approach

The genetic diversity and potential adaptability primarily extend the capabilities of GAs to solve complex real-world problems belonging to financial market forecasting. In this article, we propose to apply information fusion to the internal operations of GA; parallel operations conducted on the two parts of given chromosomes are enhanced with inter as well as intra-chromosome crossover along with adaptive mutation. Such a fusion can introduce a higher level of genetic diversity to improve the search ability of potential solutions. A structural overview of the proposed information fusion-based GA approach with ICAN for stock trend prediction is given by Fig. 1; we broadly combine a series of operations under pre-processing, preparation, parameter optimization and feature selection, and prediction task blocks.

3.1.1. Time-series data collection

One of the initial steps in our proposed approach is to collect the time-series data. For an unbiased evaluation of our approach, we consider the stock data as specified in the referred articles [16, 17]. We download daily stock price data of Korea Stock Price Index (KOSPI), as well as six individual companies namely, Samsung Electronics Co. Ltd. (005930.KS), SK Hynix Inc. (000660.KS), Hyundai Motor Co. (005380.KS), POSCO (005490.KS), SK Telecom Co. Ltd. (017670.KS), and Hyundai Mobis Co. Ltd. (012330.KS) for the duration $01-01-2000$ to $31-12-2016$ [16] from Yahoo finance website [38]; we also download historical stock price data of China construction bank (0939.HK) and CSI 300 (3188.HK) for the duration $01-01-2010$ to $01-04-2020$ [17]. The raw form of such data consists of the historical stock market information for the given trading day such as the opening and the closing prices, the highest and the lowest prices attained, as well as the volume information; the same is denoted as open (O), close (C), high (H), low (L), and volume (V), respectively.

3.1.2. Cleaning

The downloaded historical time-series data may have incomplete or missing values; the unavailability of such information can lead to error propagation. Therefore, it is essential to apply a cleaning-based pre-processing step in order to clean the raw data. Here, the data rows containing any missing value(s) or 'null' value(s) are eliminated from being considered for additional feature calculation. This, in turn, ensures that possible errors are not introduced due to empty data samples.

3.1.3. Additional features

Along with the raw information collected from the given dataset, several features can play an important role in determining the market characteristics. Such features can be derived from the existing information or appended from other associated resources; technical indicators are widely adopted as the informative set of features that can provide technical assistance in studying financial markets. In our proposed approach, we integrate the parameter optimization and feature selection capabilities of GAs from the reference articles [16,17], respectively. For a fair comparison, we derive five technical indicators as considered in Ref. [16] viz. simple moving average (SMA), weighted moving average (WMA), relative strength index (RSI), stochastic %K, and stochastic %D. Subsequently, for an unbiased comparison with Ref. [17], we derive SMA, exponential moving average (EMA), moving average convergence/divergence (MACD), stochastic oscillator (OSC), and on balance volume (OBV) features. These indicators are calculated from the cleaned raw input data and they serve as additional features for our proposed approach. Hence, a total of 10 features are considered for the experimentations, however, depending on the application requirements, other features can also be derived and appended.

3.1.4. Normalization

After applying cleaning to the raw data and deriving the additional features, it can be observed that different features can have different data ranges. For example, the open price of KOSPI was found to be $2,976.81$ whereas its volume was 969,300 on the trading day $02-01-2021$ [39]. In such a scenario, features with a larger data range can dominate features with smaller data ranges; this can also affect the model's learning phase, resulting in a biased prediction. In order to ensure that equal importance is given to all the considered features, data normalization is applied to each feature using Eq. (1) [40].

$$X' = (b - a) \cdot \frac{(X - X_{min})}{X_{max} - X_{min}} + a \tag{1}$$

where, $[X_{min}, X_{max}]$ presents the original data range and each feature of the dataset is transformed to the range $[a, b]$ with a and b indicating two arbitrary points; X' denotes the normalized value of a feature value X [40]. Here, each feature is individually normalized within [0, 1] range.
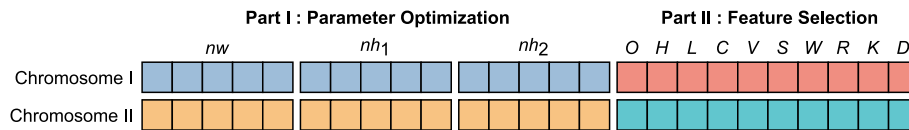
Fig. 2. An illustration of two selected chromosomes from the initial population with individual parts and their representations.

3.1.5. Data samples

The pre-processed data samples can be further prepared for training, validation, and testing; while validation can be optional, we consider the data sample division ratios as given in the referred articles [16,17]. The data are divided into 65%, 15%, and 20% for training, validation, and testing, respectively, while comparing our proposed approach with that of Ref. [16]. On the other hand, the data are divided into training and testing with 80% and 20%, respectively, while comparing our proposed approach with that of Ref. [17]. The training, as well as validation, data samples are provided to the parameter optimization and feature selection module as shown in Fig. 1 whereas the performance is evaluated using the testing data samples through the prediction module.

3.1.6. Population

The module containing parameter optimization and feature selection operations includes our proposed GA with ICAN approach where inter-intra crossover and adaptive mutation are highlighted in separate blocks in Fig. 1. The initial stage of GA is to generate a population of chromosomes that are potential solutions to the given problem. We adopt the problems of parameter optimization from Ref. [16] and feature selection from Ref. [17]; we further append these two problems within a single chromosome such that a chromosome would contain two parts as shown in Fig. 2.

Based on the reference article [16], we aim to optimize three parameters, i.e., window size (nw), number of nodes in hidden layer 1 of LSTM (nh_1), and number of nodes in hidden layer 2 of LSTM (nh_2) [16], such that their values between 1 and 31 are represented in binary with five gene positions devoted to each parameter; this results into a total of 15 genes belonging to the first part, i.e., parameter optimization. Subsequently, we target to select a set of features from the given list of 10 features viz. open (O), high (H), low (L), close (C), volume (V), SMA (S), WMA (W), RSI (R), stochastic %K (K), and stochastic %D (D) [16]. Therefore, the second part, i.e., feature selection, consists of a total of 10 genes with binary values such that a '0' would indicate elimination of the specific feature whereas a '1' would denote a selection of the specific feature while considering the given chromosome for experimentation. Hence, a total of 25 genes are combined together to represent each chromosome.

On the other hand, the reference article [17] has a window size of 5 and a single hidden layer of LSTM with 128 neurons. To represent the same in binary, three gene positions are devoted to nw parameter such that the range is [1, 5] and eight gene positions are devoted to nh_1 parameter such that the range is [1, 128]; for a fair comparison with Ref. [17], the second hidden layer of LSTM is not included and hence, nh_2 parameter is not considered in case of evaluating Ref. [17]. Hence, a total of 11 genes belong to the first part, i.e., parameter optimization. The second part aims to select a set of features; for this purpose, open, high, low, close, volume, SMA, EMA, MACD, OSC, and OBV features are considered from Ref. [17]; therefore, 10 genes are devoted to the second part, similar to the case of representing features from Ref. [16] for our proposed approach. Hence, each chromosome of our proposed approach consists of a total of 21 genes while being compared with the existing approach [17] and corresponding changes are reflected in the chromosome representation. It must be noted that the number of genes can subsequently vary for

a given problem. These chromosomes are generated randomly in the initial population generation to create a pool of potential solutions.

3.1.7. Selection

In GA, selection of the fittest chromosome is an important factor. A large number of chromosomes are available in the initial set of population, however, selection of a set of fittest chromosomes is carried out based on the evaluation criterion. For our approach, the potential solutions are evaluated against the training (and validation) data samples such that their performance is evaluated by the considered fitness function. In other words, the gene values in each chromosome are identified to derive parameter values for nw, nh_1, and nh_2 along with a set of features; the features are selected and given as input to the prediction model whereas the parameter values are carried forward with the considered prediction model, i.e., LSTM. The performance of the same is evaluated using the identified fitness function, i.e., mean squared error (MSE); in this case, the chromosome with the minimum error shall be considered the fittest. After repeating the same with each chromosome, they are sorted using their fitness values and the decided number of fittest chromosomes are selected as parents.

3.1.8. Inter-intra crossover

The conceptual idea behind a crossover operation is to exchange a set of genes between the selected pair of parent chromosomes in order to generate a new pair of offsprings. In this article, we propose to apply an inter-intra chromosome crossover after the selection of a set of fittest chromosomes based on their fitness values. As provided by Fig. 2, each chromosome consists of two parts representing parameters and features; while we address two problems of optimization and selection, respectively, the operations on individual chromosome parts can be divided into two tasks: inter-chromosome crossover and intra-chromosome crossover.

The inter-chromosome crossover is the basic way of applying crossover between two chromosomes such that from a given point, the chromosome genes of both the parents are separated and interchanged with that of each other. While we propose to combine two target problems, i.e., parameter optimization and feature selection within a single chromosome, we apply inter-chromosome crossover on individual parts of the parent chromosomes. In other words, part I and part II of each parent chromosome are randomly divided and corresponding genes are interchanged with that of the other parent chromosome. Hence, two separate crossover operations are applied under inter-chromosome crossover. For a fair comparison with the existing approaches, individual crossover rates are chosen for part I and part II as given by Ref. [16,17], respectively. Also, our experiments are based on the defined target problems with a specific number of parameters to be optimized as well as features to be selected, and therefore, the length of each part of the chromosome is derived to meet the requirements. For other potential applications, the chromosome length can be varied accordingly.

Here, we propose a conditional intra-chromosome crossover operation with an aim to integrate diversity in GA operations through information fusion. The primary motivation is to diversify chromosomes in such a way that the problem of being

trapped into local optima could be addressed. For this purpose, we provide two conditions such that when one of these conditions is met, an intra-chromosome crossover is applied to one of the chromosomes of the selected parent chromosome pair, followed by a regular inter-chromosome crossover operation. The conditions (C_1 and C_2) are defined as follows: C_1 checks whether the number of generations, i.e., iterations, has reached a given threshold; here, we set the threshold to be halfway such that if half of the total number of generations are created, the intra-chromosome crossover takes place; this condition ensures that a conditional diversity is introduced at least once during the training. On the other hand, C_2 evaluates whether the fitness value of the fittest chromosome has increased within the previous generations; here, the mutation rates for each part of the chromosome are adaptively updated as explained in the following section; if the fitness of the population is not increased for consecutive two generations [41], an intra-chromosome crossover operation is carried out; this condition aims to identify situations such as being trapped into local optima and ensures that conditional diversity is introduced to provide an opportunity to explore the search space. Meeting one of these conditions sends a signal to apply intra-chromosome crossover on one of the selected parents of each chromosome pair. An intra-chromosome crossover operation indicates that genes of two parts of the same chromosome are interchanged with each other, followed by an inter-chromosome crossover operation that results in information fusion.

For evaluating our proposed approach, we adopt simulation specifications from the existing approaches [16,17]. It is possible that the number of genes in each part of the chromosome may be different. In comparison with Ref. [16], we have 15 genes in part I and 10 genes in part II; thus, the application of an interchange due to intra-chromosome crossover would replace only 10 genes of part I whereas the other 5 genes would remain unchanged. In contrast to that, when we have 11 genes in part I and 10 genes in part II [17], only one gene of part I would remain unchanged whereas the rest of the genes would be interchanged with that of part II. It must be noted that applying intra-chromosome crossover to both the chromosomes in the given pair is not desirable because the updated chromosomes further undergo regular inter-chromosome crossover operation to generate new offsprings. Here, inter-intra crossover operations are aimed to enhance exploration of the search space.

3.1.9. Adaptive mutation

In GA, mutation can be understood as a random modification to the gene(s) in order to create exploitation. Such random changes are likely to be controlled with a lower mutation rate such that only a few gene values are modified. In our approach, each chromosome has two parts and similar to the individual application of crossover rates, we propose to apply different mutation rates (as given by Ref. [16,17]) to both the parts as demonstrated in Fig. 1. Here, for simplicity of understanding, a specific color code is followed to represent part I and part II of a chromosome under crossover, as well as mutation, operations.

With an aim to enhance the optimization capabilities of GA, we propose to update the considered mutation rates in an adaptive manner. For this purpose, we store the mutation rates applied to individual parts of a chromosome in the previous generation(s); we also track fitness values attained in each generation and evaluate whether the fitness value has improved. In case of having a higher fitness value in the last generation than that of the previous generation, we increase the mutation rates; this increment is in the direction to attain the maximum accuracy and therefore, the increment is evaluated using Eq. (2).

$$increment = \frac{100 \times m_g}{f_g - f_{g-1}} \quad (2)$$

where, m_g and f_g denote mutation rate and fitness value for the gth generation. Thus, in the next generation, a mutation rate is added with the *increment* value and applied to the corresponding part of the chromosome. This process is repeated with each generation until any mutation rate reaches the upper bound, i.e., $m_g = 1$. In the case of $m_g \geq 1$, the mutation rate of the previous generation is considered. On the other hand, if the fitness value gets decreased even after having an adaptive increase in the mutation rates, the rate values are decreased accordingly; subsequently, if no modification is found in the fitness value of the fittest chromosome in two consecutive generations, an intra-chromosome crossover operation is carried out under condition, C_2. While adaptive mutation enhances exploitation, its operations are also supported by conditional intra-chromosome crossover to improve exploration of the search space. Together, these operations target to optimize the given set of problems by introducing genetic diversity. A step-by-step procedure to apply inter-intra crossover and adaptive mutation is provided in Algorithm 1. Here, part I and part II of each of the selected two chromosomes are presented to have n_1 and n_2 number of genes, respectively; while we have $n_1 > n_2$ in the considered existing approaches [16,17], our generalized algorithm can be easily adapted for any other combination of a number of genes in each part of a chromosome.

Algorithm 1 Inter-intra crossover and adaptive mutation

    Input: Two chromosomes, Ch_1 and Ch_2 with n_1 and n_2 numbers of genes in part I and part II, respectively; current generation, g; total generations, G; increase in fitness, $f_{increment}$; mutation rate, m_g, conditions, $C_1 = 0$ and $C_2 = 0$

    Output: Two updated chromosomes

1: **if** $g == \frac{G}{2}$ **then**
2:     Condition, $C_1 = 1$
3: **end if**
4: **if** $f_{increment} == 0$ **then**
5:     Condition, $C_2 = 1$
6: **end if**
7: **if** $C_1 == 1$ or $C_2 == 1$ **then**
8:     Select $n = min(n_1, n_2)$ number of genes in each part of Ch_1
9:     Swap the selected n genes of each part in Ch_1
10: **end if**
11: Select two random points, r_1 and r_2 in part I and part II, respectively
12: Apply crossover with individual parts of Ch_1 and Ch_2 using r_1 and r_2
13: Calculate *increment* for each part using eq. (2)
14: **if** *increment* > 0 **then**
15:     Increase m_g by amount of *increment* for each part
16: **else if** *increment* < 0 **then**
17:     Decrease m_g by amount of *increment* for each part
18: **end if**
19: **if** $m_g \geq 1$ or $m_g \leq 0$ **then**
20:     Select $m_g = m_{g-1}$
21: **end if**
22: Apply mutation on each part of Ch_1 and Ch_2 with respective mutation rates
23: **return** Updated chromosomes

3.1.10. Prediction model

While the proposed GA with ICAN approach aims to optimize the network parameters and select a set of optimal features, the evaluation of this approach is carried out using an LSTM prediction model. As considered by the referred article [16], we adopt LSTM and determine parameters nw, nh_1, and nh_2, i.e., window size, number of nodes in the hidden layer 1, and number

of nodes in the hidden layer 2, respectively; we also use the chromosome to represent these parameters along with the considered 10 features. Subsequently, comparison with the referred article [17] requires only two parameters viz. nw and nh_1. For each generation, the potential GA with ICAN-based solutions are evaluated with the set of features as inputs and derived set of parameters using the data samples as described earlier. An LSTM cell consists of state (C_t), forget gate (f_t), input gate (i_t), output gate (o_t), and input vector (x_t); the internal operations of the same can be given by Eqs. (3) to (8) [42].

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * tanh(C_t) \tag{8}$$

where, weight matrices can be denoted using W_f, W_i, W_C, and W_o whereas bias vectors are given by b_f, b_i, b_C, and b_o; σ, i.e., sigmoid and $tanh$ indicate the activation functions; $\tilde{C}_t$ is a candidate cell state and h_t defines the output vector at time t. For each generation, the chromosome fitness is evaluated using this prediction model with training (and validation) data samples; also, when the termination criterion is met, testing data samples are fed to the same prediction model in order to derive the prediction performance.

3.1.11. Fitness function

One of the critical aspects in GA is a fitness function; it evaluates the suitability of the chromosome in deriving a potential solution to the given problem. Here, the chromosomes represent parameters to be optimized and features to be selected. Therefore, the application of such combinations to the prediction model decides how capable corresponding chromosome solutions are in obtaining useful solutions. Among various fitness functions, MSE is selected as a fitness function for LSTM prediction model in the considered reference article [16] as given by Eq. (9).

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \tag{9}$$

where, Y_i and $\hat{Y}_i$ denote actual output and predicted output for the ith record, respectively; N presents the number of data samples. On the other hand, R^2 score, also known as coefficient of determination, is adopted as a fitness function in the referred article [17] which can be calculated using Eq. (10).

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{Y}_i)^2} \tag{10}$$

where, Y_i and $\hat{Y}_i$ indicate actual output and predicted output for the ith record, respectively; $\bar{Y}_i$ denotes mean of Y_i; N is the total number of data samples. As our proposed approach consists of two parts in chromosome for parameter optimization and feature selection, we adopt specific crossover rates and mutation rates for individual parts; in continuation to the same, we adopt MSE for part I [16] and R^2 score for part II [17]. Here, with each chromosome, individual fitness values of both the parts are calculated and the fittest chromosome(s) with smaller MSE and higher R^2 score is selected. Here, it is found that these metrics operate as complements of each other; in other words, with a decreasing

MSE, an increase in R^2 score can be observed, and vice-versa. We consider stock price trend prediction problem wherein MSE metric is aimed to be decreased; thus, for simplicity of the final optimization calculations, MSE fitness function is considered to minimize the given optimization problem.

3.1.12. Optimized parameters

Based on chromosome fitness, a set of fittest chromosomes is derived; these chromosomes serve as potential solutions and are carried forward to the next generation of the population for further evaluation until the termination criterion is met. When the termination criterion, for example, the maximum number of generations, is met, the optimized parameters are derived from the fittest chromosome and the same are given to the prediction module. Thus, from the fittest chromosome, a set of features, as well as a set of optimized parameters, are identified and the same are utilized with the testing data samples and prediction model, respectively. The generation-wise operations of the proposed GA with ICAN can be visualized using a flowchart as shown in Fig. 3. It represents selection operation, followed by conditional intra-chromosome crossover and inter-chromosome crossover operations; further, conditions are set in the proposed approach to demonstrate adaptive mutation operations. Such operations are carried out in each generation and the fitness value of each chromosome is calculated; these values indicate the optimized parameters in part I and the selected set of features in part II of a chromosome. Hence, the population of a new generation is created and these steps get repeated until the termination criterion, e.g., the maximum number of generations, is met.

3.1.13. Stock price and trend prediction

The proposed approach based on GA with ICAN aims to predict one-day-ahead stock close price and trend for the targeted companies. For this purpose, the selected features and optimized parameters are provided to forecast the stock price and the price movement, i.e., stock trend is further derived. In order to determine the performance of the prediction, various evaluation metrics are taken into consideration as the next step.

3.1.14. Performance evaluation

The stock price prediction can be further evaluated for stock trend forecasting. Therefore, the prediction capability of our proposed approach is evaluated using three performance metrics viz. MSE, mean absolute error (MAE), and mean absolute percentage error (MAPE) as given by Eq. (9), (11), and (12), respectively.

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^{N} |Y_i - \hat{Y}_i| \tag{11}$$

$$MAPE = \frac{1}{N} \cdot \sum_{i=1}^{N} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \tag{12}$$

where, Y_i and $\hat{Y}_i$ denote the actual and the predicted outputs of the ith record, respectively; N is the total number of data samples. These metrics determine the errors in the predicted price values; the smaller error indicates higher prediction performance and thus, an improved stock trend prediction which is evaluated using R^2 score given by Eq. (10).

3.2. An illustrative example

The proposed ICAN approach is a combination of inter and intra-chromosome crossover operations along with adaptive mutation. For a clear understanding of the proposed work, we provide an illustrative example by considering two chromosomes and demonstrate their journey through the steps followed by
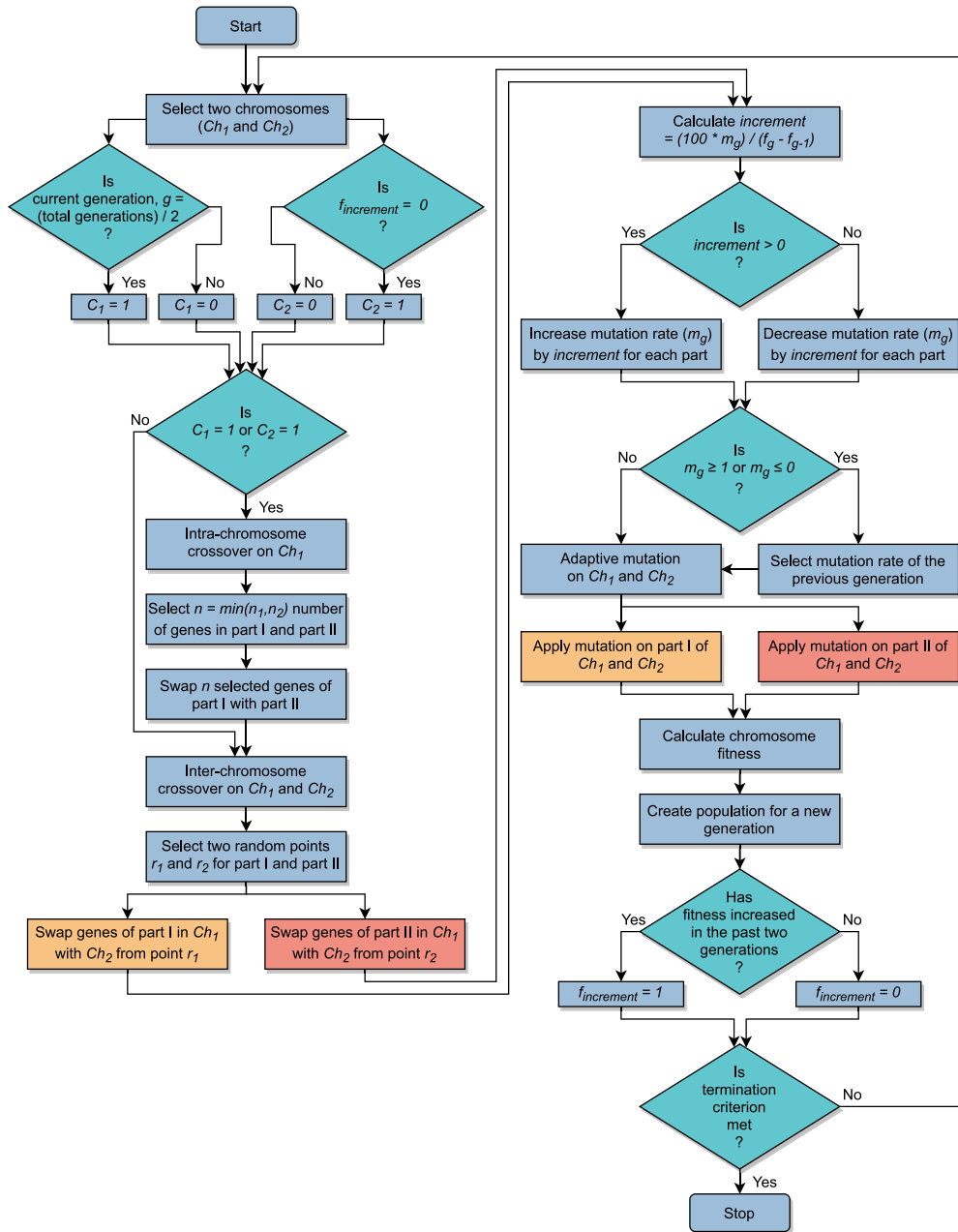
Fig. 3. Flowchart of GA with ICAN operations.

ICAN. Figs. 2 and 4 to 10 are the graphical representations of an example developed for illustration; here, the colors are specifically chosen to describe individual parts of each chromosome and the same process is carried forward throughout this illustrative example.

Fig. 2 denotes two chromosomes (Chromosome I and Chromosome II) selected from a pool of the initial population based on their fitness values; here, each chromosome consists of 25 genes divided into two parts viz. parameter optimization (Part I with 15 genes) and feature selection (Part II with 10 genes). Here, the vertically separated sets of chromosome genes are only for the demonstration purpose, however, a binary string of the length as specified by the number of genes is provided as a potential solution during the actual implementation. For the remaining steps of our illustrative example, each chromosome and its offsprings (Offspring I and Offspring II) are presented with the above-defined two parts. It must be noted that the

number of genes and corresponding representations in each part given by Fig. 2 are the adaptation of the considered existing approaches [16,17] for a fair comparison, however, the same can be modified accordingly in case of having a different target such as the parameters to be tuned and/or the pool from which features are to be selected.

The next step of our proposed approach indicates an inter-chromosome crossover applied to the individual parts of two selected chromosomes in the current generation as shown in Fig. 4; here, single-point crossover operations are applied to individual parts, denoted with two verticle dashed lines. It means the two chromosomes interchange their genes from the defined verticle dashed line in individual parts and generate new offsprings. The same can be visualized from Fig. 4 with the change in genes in each offspring. Here, the crossover is applied with individual crossover rates for parameter optimization and feature selection parts as defined by the corresponding reference articles [16,17].
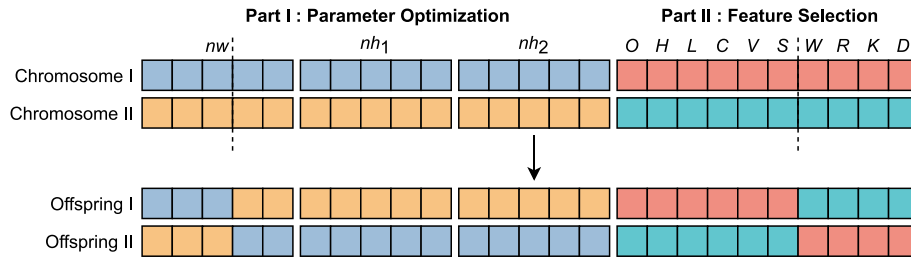
Fig. 4. An illustration of inter-chromosome crossover applied to the individual parts of two selected chromosomes in the current generation.
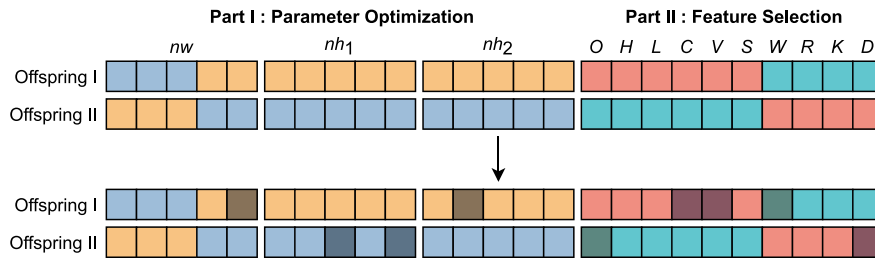

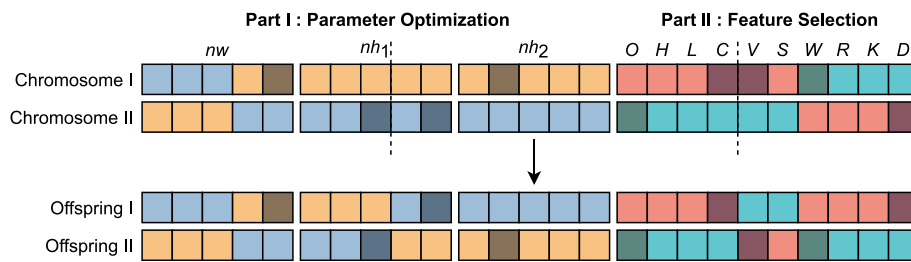**Fig. 5.** An illustration of mutation with initial mutation rates applied to the individual parts.

Fig. 6. An illustration of inter-chromosome crossover applied to the individual parts of two selected chromosomes in the next generation.

After crossover operation, the individual parts of the offsprings undergo mutation operation; Fig. 5 indicates the mutated gene position with bit-flip mutation, presented in a darker color shade than that of the actual gene position color. It must be noted that the initial mutation rates for individual parts are different, as defined by the reference articles [16,17]. At this stage, the mutated offsprings are evaluated using the fitness function and the fittest chromosomes are selected for the next generation. In our case, the parameter values, as well as the set of features, are provided to the prediction model, LSTM, and the MSE value is calculated for the predicted stock prices. For simplicity of the ongoing example, we assume that the illustrated offsprings (as shown in Fig. 5) are further given as chromosomes to the next generation.

In the next generation, a similar process of inter-chromosome crossover is carried out as shown in Fig. 6. Here as well, the crossover operations are applied to the individual parts of the selected chromosomes with the defined individual crossover rates and new offsprings are generated; the crossover points are displayed using verticle dashed lines in Fig. 6. After the first generation, ICAN proposes to update the mutation rate adaptively; hence, at this point, the new mutation rates are calculated with respect to the percentage of improvement attained in the previous generation. As given by Fig. 7, the genes of individual offspring parts get mutated with updated mutation rates; such mutated gene positions are illustrated with an alternate color shade. In other words, if a gene position consisted of a lighter color shade, its mutation would be shown with a darker color shade, and vice-versa. As compared to the number of mutated

gene positions using initial mutation rates (Fig. 5), the updated mutation rates are shown to have a higher number of gene positions mutated in the next generation (Fig. 7) which means the mutation rates are adaptively increased in that specific generation. In contrast to that, in case the mutation rates were decreased, a comparatively lesser number of gene positions would have been mutated. This process of selection, inter-chromosome crossover, and adaptive mutation are repeated until the defined criterion is met; this criterion indicates the generation into which one of the selected chromosomes undergoes intra-chromosome crossover as proposed by ICAN.

We demonstrate the intra-chromosome crossover within one of the selected chromosomes of the next generation using Fig. 8; here, we select Chromosome I to illustrate the internal swapping of chromosomes. As the numbers of genes in the considered two parts are different, i.e., 15 and 10, we showcase how the first 10 genes of Part I get interchanged with the 10 genes of Part II. One of the practices followed by the proposed approach is to have a crossover of the number of genes of individual parts. The updated chromosomes are further given to an inter-chromosome crossover operation for information fusion as shown in Fig. 9, followed by the mutation operation based on the updated mutation rates as shown in Fig. 10. With the increasing number of generations, intra-chromosome crossover can take place several times, depending on the criterion. When the termination criterion is met, parameter values and a set of features are taken from the fittest chromosome and their prediction performances are evaluated.
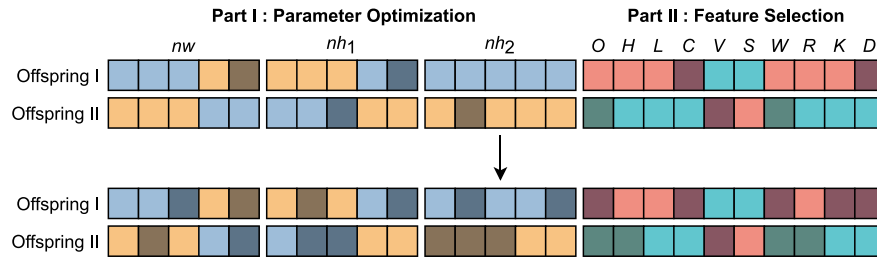
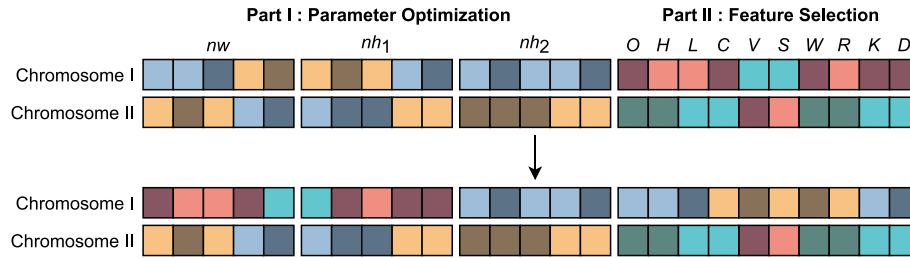**Fig. 7.** An illustration of adaptive mutation with updated mutation rates applied to the individual parts.

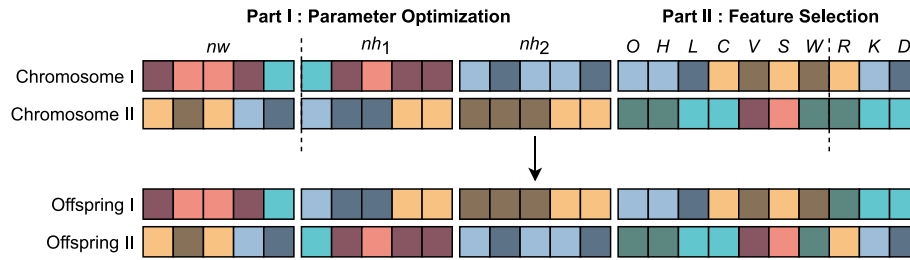**Fig. 8.** An illustration of intra-chromosome crossover applied to one of the selected chromosomes in the next generation.

Fig. 9. An illustration of inter-chromosome crossover applied to the individual parts of two selected and updated chromosomes.
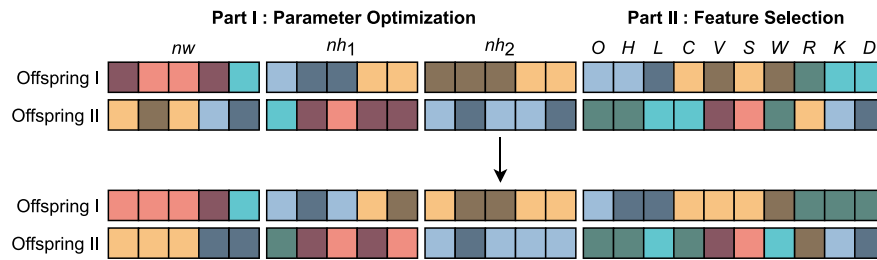

Fig. 10. An illustration of adaptive mutation with updated mutation rates applied to the individual parts.

4. Experimental analysis

In this article, we propose an information fusion approach by introducing genetic diversity to enhance the search capability of chromosomes and hence, to improve the fitness values of the potential solutions. Our proposed GA with ICAN approach introduces a series of parallel operations such that each chromosome is designed to solve two problems: parameter optimization and feature selection. The chromosome operations such as crossover and mutation, as well as fitness function, are separately adopted from the considered reference articles [16,17]; this section provides details on the experimental setup, result analysis, and a discussion on the same.

4.1. Experiments setup

In this article, we propose GA with ICAN approach to address parameter optimization and feature selection tasks; Table 1 provides detailed specifications on the set of parameters used for our proposed approach. These parameter values are mainly adopted from the reference articles [16,17] for a fair comparison. As given by Table 1, three parameters, nw, nh_1, and nh_2, are targeted to be optimized for the considered prediction model in part I of the chromosome; also, a set of features from the given list of 10 input features is to be selected by part II of the chromosome. In the referred article [16], no explicit information has been found on the range of these three parameters; the optimal values obtained by these parameters in Ref. [16] are $nw = 10$, $nh_1 = 15$, and $nh_2 = 7$. While having a zero value to any parameter may not be

Table 1
Parameter specifications for the proposed approach.

Parameter		Value [16,17]
Prediction frequency		Daily
Datasets	(1)	Korea Stock Price Index (KOSPI) [16]
	(2)	SK Hynix Inc. (000660.KS) [16]
	(3)	Hyundai Motor Co. (005380.KS) [16]
	(4)	POSCO (005490.KS) [16]
	(5)	Samsung Electronics Co. Ltd. (005930.KS) [16]
	(6)	Hyundai Mobis Co. Ltd. (012330.KS) [16]
	(7)	SK Telecom Co. Ltd. (017670.KS) [16]
	(8)	China construction bank (0939.HK) [17]
	(9)	CSI 300 (3188.HK) [17]
Data specification	(1)	$01 - 01 - 2000$ to $31 - 12 - 2016$ [16]
	(2)	$01 - 01 - 2010$ to $01 - 04 - 2020$ [17]
Normalization		[0, 1]
Input features	(1)	Open, High, Low, Close, Volume, SMA, WMA, RSI, Stochastic %K, Stochastic %D [16]
	(2)	Open, High, Low, Close, Volume, SMA, EMA, MACD, OSC, OBV [17]
Chromosome		Binary (randomly initialized)
Population size	(1)	70 [16]
	(2)	100 [17]
Crossover		Single-point crossover
Crossover rate		0.7 for part I; 0.8 for part II
Mutation		Bit-flip mutation
Mutation rate		0.15 for part I; 0.003 for part II
Number of generations	(1)	10 [16]
	(2)	100 [17]
Fitness function		MSE for part I; R^2 score for part II
Window size		nw
Prediction model (architecture)	(1)	LSTM $(10 - nh_1 - nh_2 - 1)$ [16]
	(2)	LSTM $(10 - nh_1 - 1)$ [17]
Activation	(1)	*tanh*, linear [16]
	(2)	ELU [17]
Loss		MSE
Optimizer		Adam
Number of epochs	(1)	10 [16]
	(2)	100 [17]
Training : Validation : Testing	(1)	65% : 15% : 20% [16]
	(2)	80% : − : 20% [17]
Performance metrics		MSE, MAE, MAPE, R^2 score

desirable to our approach, we set a range of [1, 31] to match the optimal values obtained by the referred article [16]. The rationale behind selecting this range is due to the best value of nh_1, i.e., 15; as we have binary chromosomes to present these parameters, at least 4 binary positions are required to determine a value as large as $2^4 - 1 = 15$. To provide at least one higher value range than 15, we require 5 binary positions to be filled in which, in turn, gives a range of 1 to 31 $(= 2^5 - 1)$ to each parameter. Thus, the presented part I of each chromosome consists of a total of 15 gene positions for three parameters. On the other hand, part II of each chromosome consists of 10 gene positions, one for each feature. With a total length of $15 + 10 = 25$ genes in each chromosome, our proposed GA with ICAN approach is performed.

On the other hand, two parameters, nw and nh_1, are aimed for optimization in part I of the chromosome for comparison with the referred article [17]. The existing article specifies $nw = 5$ and $nh_1 = 128$; as explained earlier, zero value to any of these two parameters is not desirable in our approach, and therefore,

we set a range using 3 and 8 binary positions, respectively, to ensure that the defined parameter values can be obtained from the chromosomes. Here, $2^3 - 1 = 7$ and $2^8 - 1 = 255$ can be the maximum attainable values for nw and nh_1, respectively. Thus, comparison of our proposed ICAN approach with Ref. [17] requires $3 + 8 = 11$ gene positions in part I and 10 gene positions in part II, and hence, a total of $11 + 10 = 21$ genes in each chromosome.

4.2. Result analysis

The proposed approach based on GA with ICAN is designed to address parameter optimization and feature selection problems. To provide a fair comparison with the existing approaches [16, 17], we initially experiment using the defined parameter values (nw, nh_1, and nh_2) and with 10 features for seven datasets as provided by Ref. [16] and for two datasets as specified by Ref. [17]. We further execute our proposed approach by providing

Table 2

Comparison between existing optimization GA [16] or GA [17], ALO, and proposed GA with ICAN approaches.

Dataset	Approach	MSE		MAE		MAPE		R^2 Score	
		Average	SD	Average	SD	Average	SD	Average	SD
KOSPI	GA [16]	178.0015	71.4886	11.9658	2.4620	0.1767	0.0379	0.0300	1.0616
	ALO	50.1863	2.2211	6.0336	3.2728	0.0890	3.1552	0.1182	0.4304
	GA with ICAN	33.9963	24.0331	4.7488	1.8663	0.0706	0.0270	0.2422	0.5331
000660.KS	GA [16]	5.3470	2.1339	1.8801	0.4570	2.9538	0.9281	0.8151	0.0764
	ALO	5.2219	1.2517	1.8273	3.8239	2.7916	5.0166	0.8152	0.0954
	GA with ICAN	3.7051	1.7022	1.5540	0.4180	2.3389	0.7388	0.8688	0.0600
005380.KS	GA [16]	123.4161	37.8568	8.2701	1.4573	0.1438	0.0154	0.9143	0.0278
	ALO	94.3945	0.8756	7.4684	4.5412	0.1289	4.4234	0.9309	0.0341
	GA with ICAN	58.5383	34.3778	5.6686	1.5226	0.1090	0.0183	0.9572	0.0252
005490.KS	GA [16]	40.3031	6.4077	5.0148	0.4358	0.7999	0.0904	0.9301	0.0116
	ALO	33.9381	1.1353	4.5621	3.9285	0.7518	3.2128	0.9389	0.0211
	GA with ICAN	26.4833	9.1796	4.0444	0.7114	0.6888	0.1211	0.9524	0.0165
005930.KS	GA [16]	877.0813	376.65	26.6943	5.7671	0.2792	0.0627	0.3125	0.6953
	ALO	305.1311	3.8930	13.3536	5.5588	0.1400	4.6619	0.4359	0.7257
	GA with ICAN	194.2573	182.0443	10.5170	5.9972	0.1112	0.0614	0.6428	0.3331
012330.KS	GA [16]	65.5324	11.2084	6.4476	0.6875	0.1002	0.0124	0.8123	0.0394
	ALO	32.2166	0.4831	4.4663	5.1651	0.0669	3.6893	0.8859	0.0400
	GA with ICAN	38.8923	12.5864	4.9888	0.9137	0.0748	0.0110	0.8624	0.0446
017670.KS	GA [16]	306.3974	75.4051	14.0204	2.3046	1.1918	0.1836	0.7869	0.0567
	ALO	182.7673	6.8313	10.7690	6.8028	0.9812	5.4772	0.8625	0.0261
	GA with ICAN	181.3421	100.7718	10.5540	3.0703	0.9299	0.1646	0.8637	0.0759
0939.HK	GA [17]	118.1713	9.5425	8.6183	0.3721	1.5132	0.1089	0.9208	0.0064
	ALO	109.6096	6.0997	8.4056	0.6566	1.4038	0.5900	0.9329	0.0091
	GA with ICAN	101.0480	14.7539	8.1929	0.5879	1.2943	0.1682	0.9342	0.0104
3188.HK	GA [17]	54.5555	17.5221	5.7644	0.9447	0.4933	0.0506	0.8511	0.0478
	ALO	50.4983	2.2211	5.5795	0.3273	0.4908	0.3155	0.8719	0.0430
	GA with ICAN	46.4412	17.8310	5.3945	1.2262	0.4884	0.1049	0.8727	0.0490

the defined parameter range and feature set using GA with ICAN to determine the optimal parameter values and a set of features. For a fair comparison, we repeat each simulation 10 times and the resultant metrics using our proposed approach are evaluated using MSE (Eq. (9)), MAE (Eq. (11)), MAPE (Eq. (12)), and R^2 score (Eq. (10)) metrics.

While existing and proposed approaches are based on an evolutionary algorithm, GA, we extend our experiments to include one of the recent advances in swarm intelligence. For this purpose, we adopt ALO to optimize the above-defined parameters and to select a set of features. The existing approaches using GA, as well as ALO, and the proposed approach using GA with ICAN are individually applied to an LSTM prediction model and an initial comparison of the predicted stock prices is conducted using three error-based performance metrics namely, MSE, MAE, and MAPE, as well as a trend deterministic R^2 score metric. Here, with respect to each metric, the comparison is carried out between the existing optimization approach using GA [16,17], ALO, and the proposed approach for respective datasets using average and standard deviation (SD) as shown in Table 2. Here, "GA [16]" and "GA [17]" denote the existing approaches whereas, "ALO" and "GA with ICAN" represent the application of respective approaches for parameter optimization and feature selection-based stock price trend prediction. It can be observed that the overall average performance of our proposed approach for stock price trend prediction is remarkably higher than the existing GA as well as ALO-based optimization approaches; also, the SD values indicate considerably stable prediction performance using the optimized parameter values as well as selected features derived by our proposed approach. As we consider comparing our proposed GA with ICAN approach with the existing GA as well as ALO-based optimization approaches on a total of nine datasets and repeat each experiment 10 times, our simulations undergo $9 \times 10 \times 3 = 270$ runs and show average and SD values for a fair comparison.

MSE, MAE, and MAPE are some of the widely integrated error determination metrics; smaller errors denote closer prediction

outputs to that of the actual output values. The comparison of the predicted stock price errors is shown in Figs. 11–13 for MSE, MAE, and MAPE, respectively, for all the nine considered datasets. Fig. 11 shows a comparison using MSE such that for each dataset, an average prediction performance shows a significant reduction in the forecasting error using our proposed GA with ICAN approach. It can be visualized that for each dataset, our proposed approach attains a remarkably lower MSE value than that of the existing approaches with GA [16,17]. We extend the analysis by incorporating ALO-based optimization; it can be viewed that the error rates using ALO are lower than that of GA, however, the proposed GA with ICAN approach is capable of attaining the smallest errors than GA as well as ALO. The results indicate a capability of GA with ICAN approach in optimizing parameters and selecting features as compared to evolutionary algorithm GA as well as swarm intelligence-based ALO. The significance of integrating genetic diversity through information fusion can be seen through other metrics; Figs. 12 and 13 denote stock price prediction comparisons using MAE and MAPE metrics, respectively. It can be observed that for each dataset of the referred articles [16,17], our proposed GA with ICAN approach reduces the error rates and thus, enhances the prediction capability of LSTM by providing a set of features and optimized parameter values. Here, ALO-based optimization results in higher error rates than that using GA with ICAN which denotes the adaptability of our proposed approach for complex stock market applications.

While we examine the error rates obtained through various experiments, it is important to analyze the resultant values that could improve the prediction performance. Here, we have various parameters (nw, nh_1, and nh_2) and a number of features to be optimized and selected. To evaluate the same, we consider R^2 score metric to determine the stock price trend prediction performance and represent the same with nw, nh_1, nh_2, and number of features as shown in Figs. 14 and 15 for comparison with the existing approach [16,17], respectively. Here as well, we demonstrate the results using optimization approaches GA, ALO,
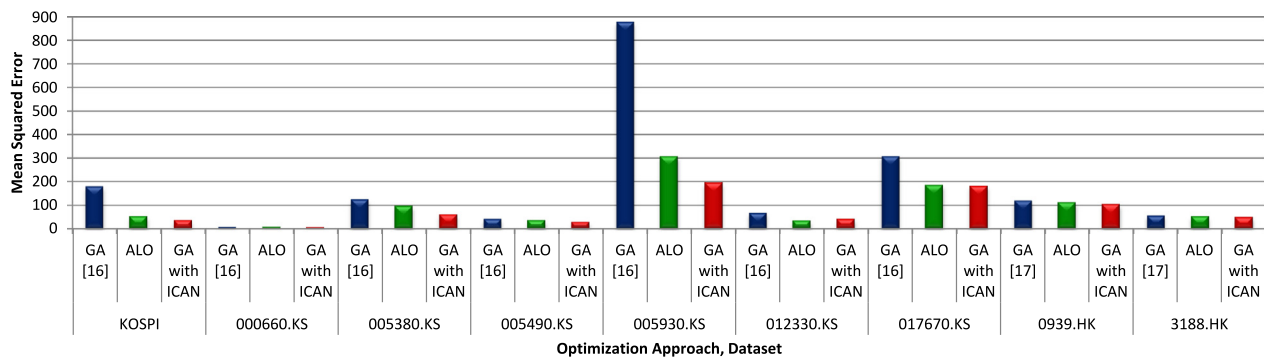
Fig. 11. Comparison between existing optimization GA [16] or GA [17], ALO, and proposed GA with ICAN approaches with various datasets using MSE.
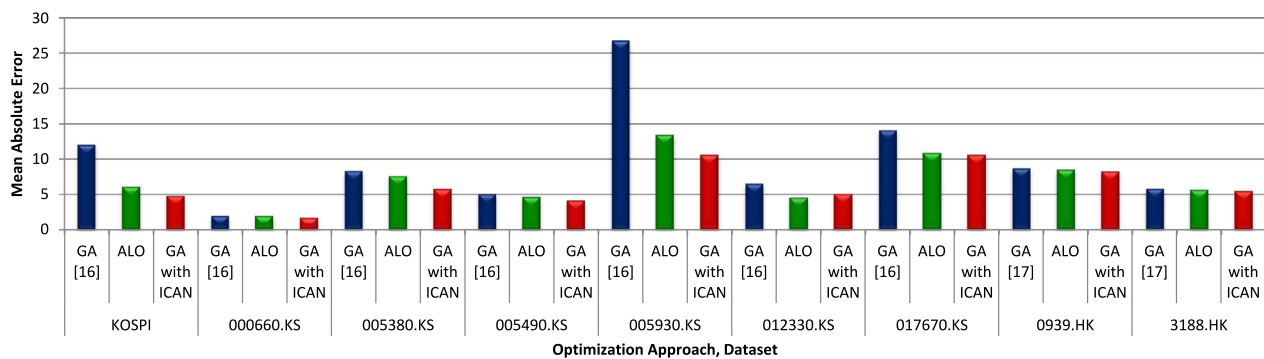

Fig. 12. Comparison between existing optimization GA [16] or GA [17], ALO, and proposed GA with ICAN approaches with various datasets using MAE.
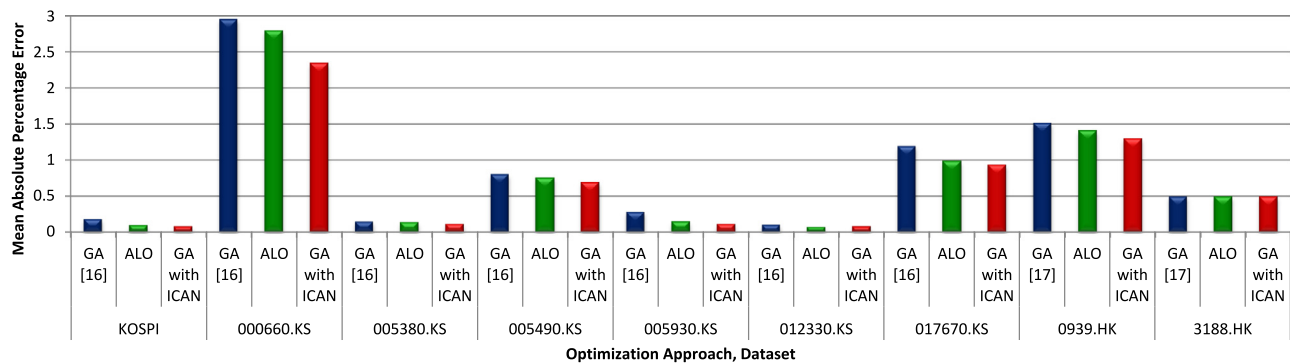

Fig. 13. Comparison between existing optimization GA [16] or GA [17], ALO, and proposed GA with ICAN approaches with various datasets using MAPE.
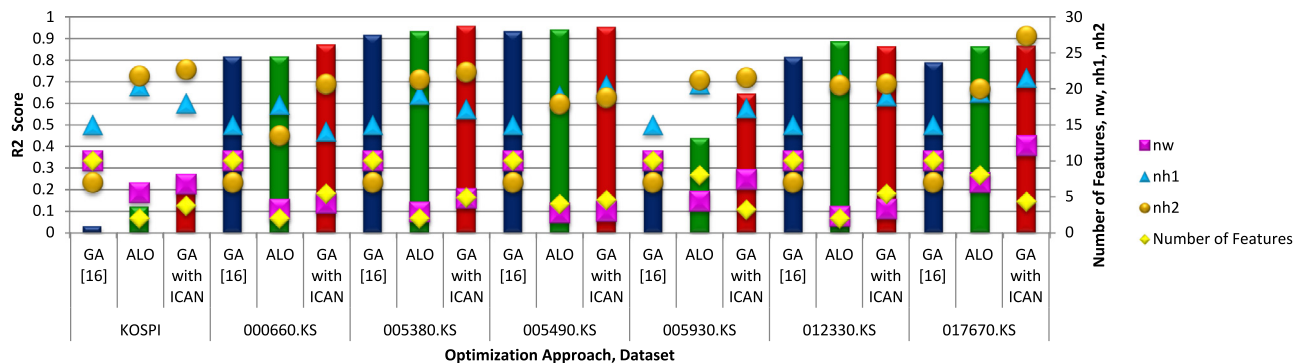

Fig. 14. Comparison of the optimized parameter values and number of features selected by existing optimization GA [16], ALO, and proposed GA with ICAN approaches with various datasets using R^2 score.
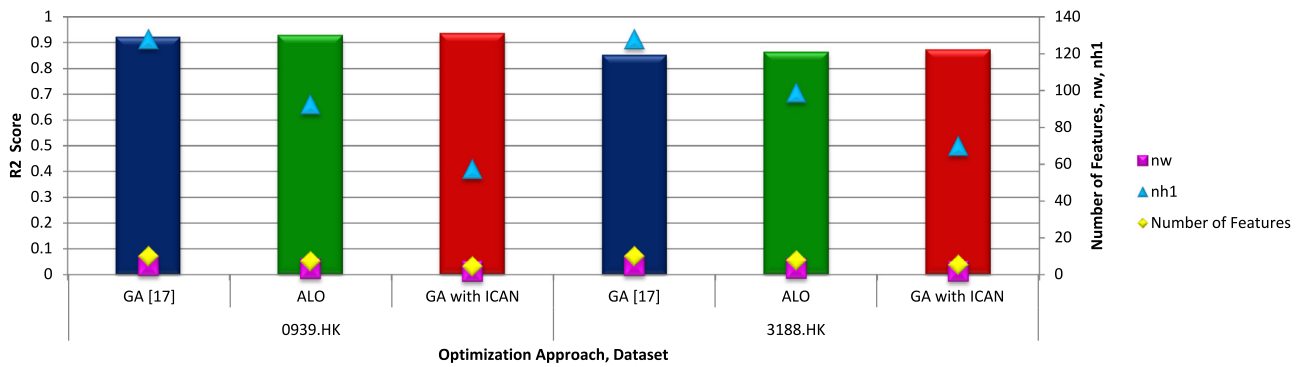
Fig. 15. Comparison of the optimized parameter values and number of features selected by existing optimization GA [17], ALO, and proposed GA with ICAN approaches with various datasets using R^2 score.

and GA with ICAN. The side-by-side bar charts indicate R^2 scores obtained by corresponding approaches; the square, triangle, and round shapes present the values of optimized parameters nw, nh_1, and nh_2, respectively, and the diamond shapes denote the number of selected features. The representations are provided in a way to evaluate the impact of optimized parameters and a selected number of features on the model performance for stock trend prediction. Here, for each dataset of both the existing approaches [16,17], ALO-based predictions attain comparable or higher R^2 scores wherein our proposed GA with ICAN-based stock trend prediction provides the highest overall R^2 scores.

The window size (nw) parameter, presented with a square shape, demonstrates that in a majority of the cases, a smaller window size has obtained a higher prediction performance. The same observation can be viewed for ALO as well as GA with ICAN-based optimization. On the other hand, the parameters showcasing the number of neurons in hidden layers of LSTM model (nh_1 and nh_2 in Fig. 14) indicate that as compared to nh_1 of the existing approach [16], a slightly higher nh_1 value is determined by our proposed approach for different datasets; subsequently, nh_2 values are consistently higher than that of the existing approach [16] for various datasets. Subsequently, Fig. 14 denotes consistently higher nh_1 as well as nh_2 values using ALO-based optimization. Thus, it can be determined that smaller window size and a higher number of hidden layer neurons are capable of attaining a considerably smaller error rate and hence, a higher prediction performance. In contrast to that, Fig. 15 shows a remarkably lower number of hidden layer neurons (nh_1) that are optimized by ALO, as well as our proposed approach, and resulted in higher R^2 scores for both the datasets. This can be viewed as a potential approach where a prediction model with a suitable number of hidden layer neurons can significantly enhance the prediction performance.

The other part of our considered approach is dedicated to feature selection. While a total of 10 features are chosen from the referred articles [16,17], part II of each chromosome selects a set of features for our proposed approach. Therefore, it is essential to evaluate the number of selected features along with the optimized parameters. Figs. 14 and 15 also illustrate comparisons based on the number of selected features using diamond shapes. Here, the number of selected features is an average of the conducted simulation rounds. It can be observed that as compared to using all 10 features [16,17], ALO, as well as GA with ICAN, determine nearly half or less number of the features. This comparison can be helpful in understanding the significance of feature selection approach; instead of using all the features, our proposed approach chooses a significantly fewer number of features, resulting in a precise set of information and hence, enhanced prediction performance. These observations on the optimized parameter values, as well as the number of selected features, can also be viewed in Figs. 11 to 13 for error metrics. The results denote improved stock price and trend prediction using our proposed information fusion-based GA with ICAN approach and hence, it can be viewed as a potential way to address multiple optimization problems using GAs; the proposed approach enhances the optimization ability of GA with an introduction to information fusion using inter-intra crossover and adaptive mutation. Genetic diversity plays an important role in enhancing the fitness of the new populations and the same can be determined from the conducted experimentations. While we analyze the average results using MSE, MAE, MAPE, and R^2 score metrics, it is important to understand the resultant data distributions and several related aspects; the following section extends the discussion on the same.

4.3. Discussion

Data distribution is important in understanding the problem of interest; while the input data may be the same for a specific experimental setup, the resultant values using each method may vary. This is one of the primary reasons to perform an experiment multiple times to avoid any potential bias in the result. Figs. 11 through 15 denote average results of 10 experiments for each approach. In this section, we further analyze the individual results and visualize them for a deeper analysis of data distributions. For this purpose, we adopt violin plot representations of the resultant values using four metrics, i.e., MSE, MAE, MAPE, and R^2 score for each of the nine datasets as shown in Figs. 16 through 24.

A violin plot is similar to a box plot along with a probability density representation of the data; it shows data distribution which, in turn, can be useful in analyzing the resultant metric for a given approach. Here, we compare the error rates, as well as R^2 scores of GA-based optimization [16,17] and GA with ICAN-based optimization using violin plots. For example, in Fig. 16, violin plots are prepared for each metric such that the left half of each plot denotes resultant data distribution using GA-based optimization whereas the right half of the same indicates resultant data distribution using GA with ICAN optimization approach applied to KOSPI dataset. It can be viewed that each plot consists of three horizontal lines to represent the upper bound, lower bound, as well as median value of the resultant data for a given optimization approach. While error rates are expected to be reduced, an R^2 score should be increased to denote performance improvement. To observe the same, the median values of MSE, MAE, and MAPE-based violin plots of GA with ICAN are desired to be comparatively lower on the vertical axis than that of GA; contrarily, the median value of R^2 score-based violin plot using GA with ICAN is expected to be higher on the vertical axis than
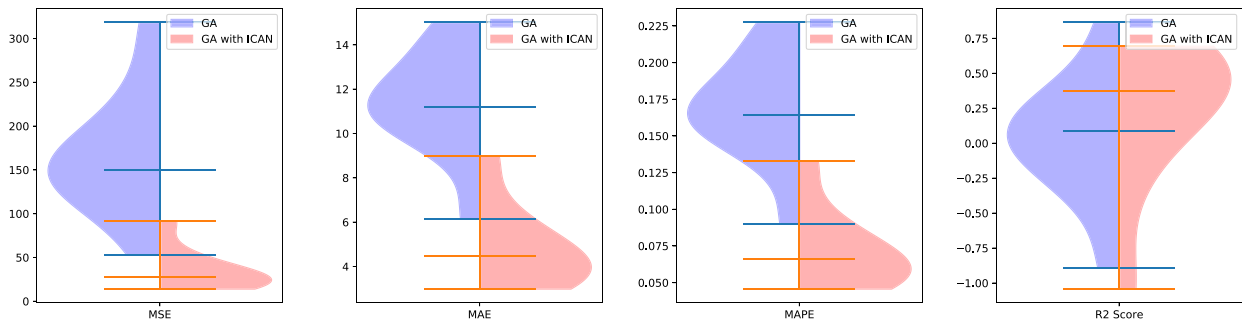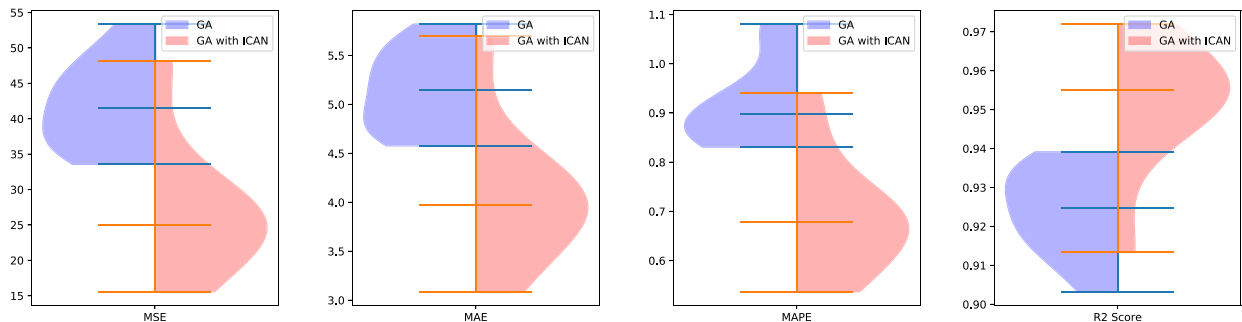
Fig. 16. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of KOSPI dataset.
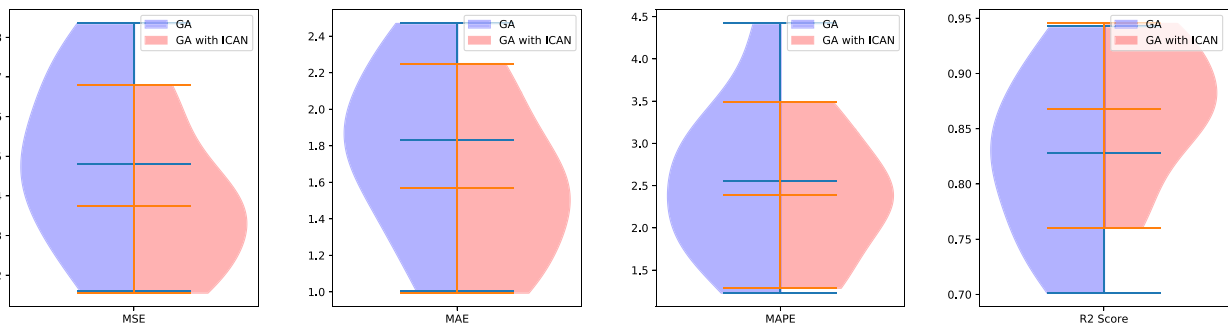
Fig. 17. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 000660.KS dataset.
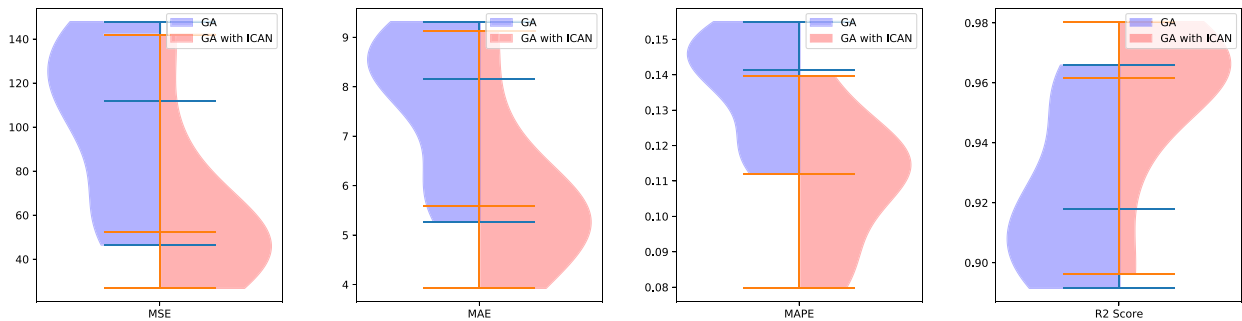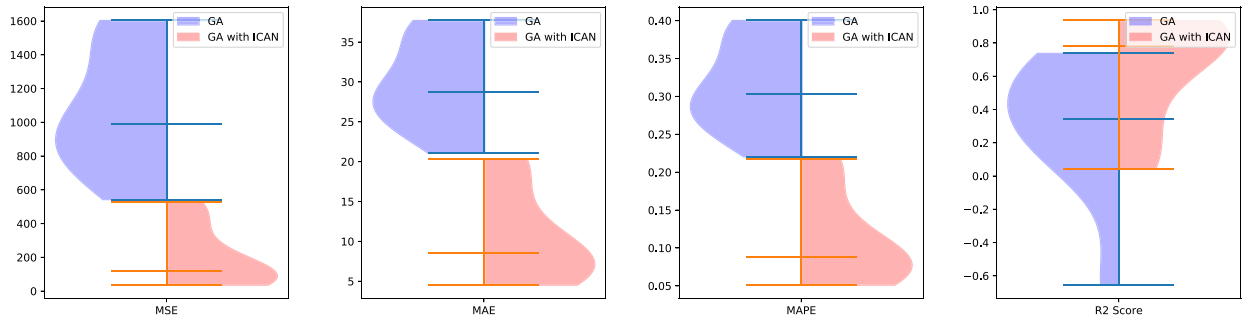

Fig. 18. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 005380.KS dataset.
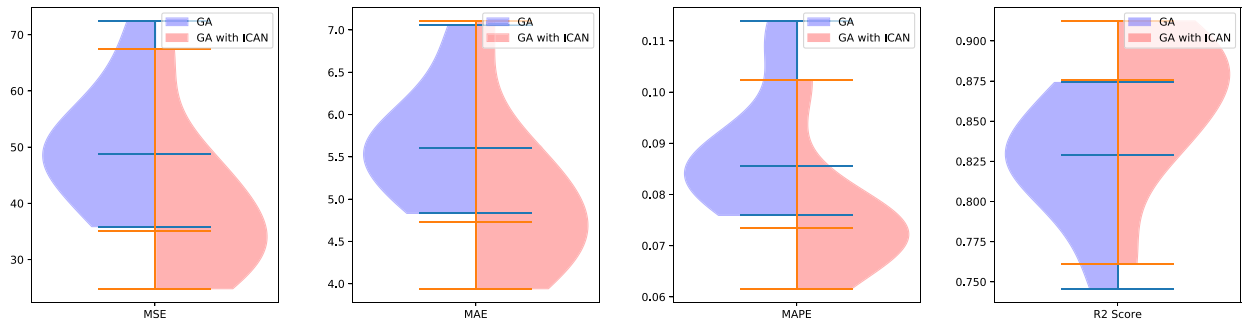

Fig. 19. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 005490.KS dataset.

that of GA. Fig. 16 denotes these features such that the resultant data distribution of the error rate-based metrics has violin plots that are vertically lower using GA with ICAN approach than that of the existing GA approach; also, the median values of GA with

ICAN can be observed to be close to the lower bounds of GA which indicate considerably lower error rates. In contrast to that, the vertically higher violin plot of R^2 score with the median value of GA with ICAN being close to the upper bound of GA
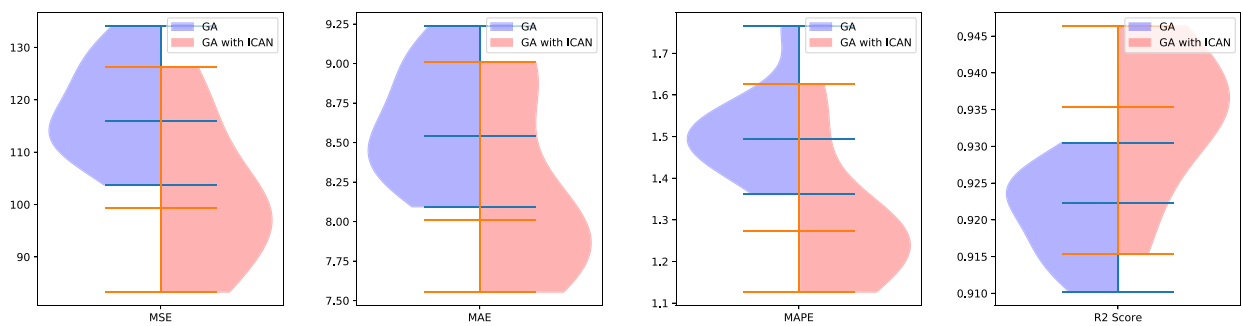
Fig. 20. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 005930.KS dataset.

Fig. 21. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 012330.KS dataset.
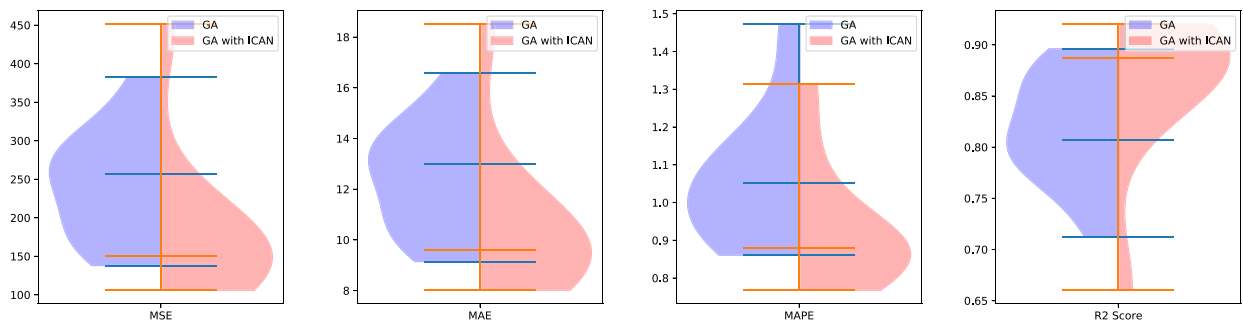
Fig. 22. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 017670.KS dataset.

Fig. 23. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 0939.HK dataset.
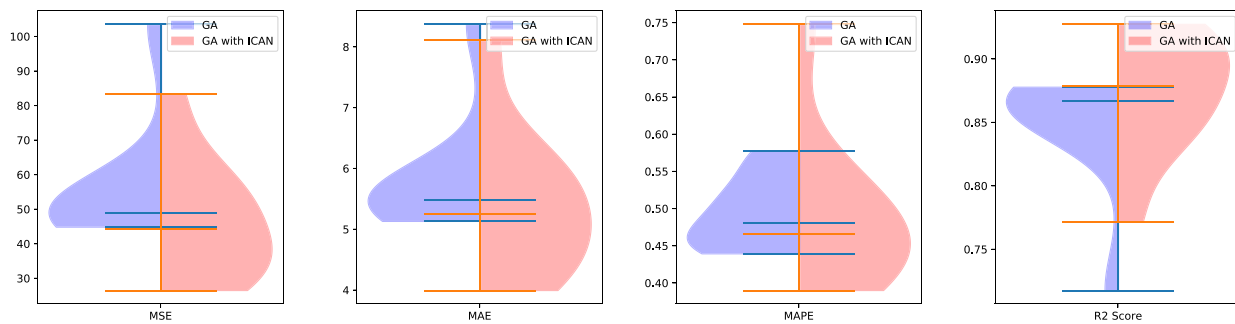
Fig. 24. Resultant data distribution in MSE, MAE, MAPE, and R^2 score metrics for GA (existing) and GA with ICAN (proposed) approaches-based optimization of 3188.HK dataset.

denote improved prediction performance. In a similar manner, violin plots derived using these metrics with other datasets can be visualized through Figs. 17 to 24.

The distribution of resultant data can explain various aspects of a prediction approach. It can be observed that certain violin plots cover a wider vertical area which indicates the spread distribution of resultant values, for example, violin plots in Fig. 17; in other words, metric results in values that cover a wider numerical range. On the other hand, several violin plots have concentrated coverage which denotes close numerical results of a given metric, for example, violin plots in Fig. 20. The considered violin plots are indicative of probability density as given by horizontal coverages. Also, in a few cases, the generated plot using GA with ICAN shows to have wider vertical coverage than that generated using GA, for example, violin plot using MAPE metric in Fig. 24; though an extended coverage means sparse data distribution, the median value obtained by the proposed GA with ICAN approach for each metric indicated remarkable performance improvement using information fusion-based GA with ICAN. Further, we evaluate the statistical significance of the derived results using Wilcoxon signed-rank test [43]; the results obtained *p*-value less than .05 for our proposed approach. Thus, it can be viewed to have a potential expansion to other related applications; also, such information fusion-based GA with ICAN approach can be potentially applied to other prediction problems. To support the same, we investigate the convergence capability of our proposed approach using two multimodal functions. As the proposed approach is based on evolutionary computation, we select test functions to be multimodal, i.e., functions that consist of multiple local optima in their search space. Such functions can explain diversification abilities as well as local optima avoidance capabilities of the targeted optimization algorithm [44,45].

In this article, we demonstrate the optimization ability of GA with ICAN and compare it with GA using Ackley and Griewank functions. The Ackley function is non-convex with a global optimum fitness of 0; similarly, the Griewank function has a global minimum fitness of 0, however, its first-order form consists of multiple maxima and minima [46]. We consider a population of 100 chromosomes with 30 genes in each chromosome that are randomly initiated with values ranging in $[-32, 32]$ for the Ackley function whereas in $[-100, 100]$ for the Griewank function as given by Ref. [45]. While crossover and mutation rates play an important role, we initiate these rates with 0.8 and 0.003, respectively, and execute GA as well as GA with ICAN approaches for 100 generations (i.e., number of iterations). The obtained fitness values for each approach are further evaluated for both the test functions as shown in Figs. 25 and 26. It can be viewed that the initial fitness of a random generation is nearly the same for GA and GA with ICAN, however, with an increase in the number of generations, our proposed GA with ICAN-based optimization approach shows commendable convergence for Ackley as well

as Griewank functions. While GA-based optimization shows no improvement in the fitness for several many generations, GA with ICAN-based optimization approach demonstrates gradual improvement in fitness after a few generations as well as a remarkable impact on the convergence speed. This can be understood from the genetic diversity perspective; as we introduce information fusion through inter-chromosome and conditional intra-chromosome crossover along with adaptive mutation operations, the approach may have higher chances of not getting stuck into local optima, and hence, improving the convergence through diversification.

For a fair comparison, we repeat the simulations using GA and GA with ICAN approaches on Ackley as well as Griewank functions for 30 times and demonstrate the data distribution in resultant fitness values using violin plots as shown in Fig. 27; here, $30 \times 2 \times 2 = 120$ simulation runs are carried out apart from the earlier 270 runs, resulting into a total of $270 + 120 = 390$ experimental runs for an unbiased evaluation of our proposed approach. For a given fitness function, we plot resultant fitness values and mark the median fitness; it must be noted that these results are obtained from 100 iterations and the convergence functions can be further evaluated to determine when a given approach reaches the global optimum. Comparison of the Ackley function indicates that our proposed approach converges faster than the existing GA approach; the median fitness value of GA with ICAN approach is considerably lower than that of the lowest fitness obtained by GA. Subsequently, fitness values obtained by our proposed approach are significantly lower than that by GA for the Griewank function; the information fusion-based GA with ICAN approach can be observed to converge remarkably faster than GA. Thus, the proposed approach can be viewed as a potential expansion of GA based on genetic diversity and the same can be used to address other complex real-world problems.

One of the limitations observed during experimental analysis is that the mutation rate of the proposed approach may suffer to remain unchanged in a long run. As described earlier, the mutation rate is adaptively modified based on the changes in fitness values; also, the previous mutation rate is adopted in case the adaptively derived mutation rate reaches beyond an upper bound. While we propose an adaptive decrement in the mutation rate in case a fitness function degrades performance, it is likely that this stage may not be arrived due to the selection and preservation of the best chromosome in each generation. For certain cases, the fitness may remain unchanged, however, the mutation rates may reach the upper bound and get rolled back to an earlier mutation rate. Such scenarios would result in the same mutation rate for the following generations. The initial mutation rate, as well as the number of generations, can play a crucial role in this regard; though such limitation may be addressed with a small initial mutation rate, the concern of an unchanged mutation rate may arise in the later generations. Hence, it can be considered a potential area of improvement in future work.
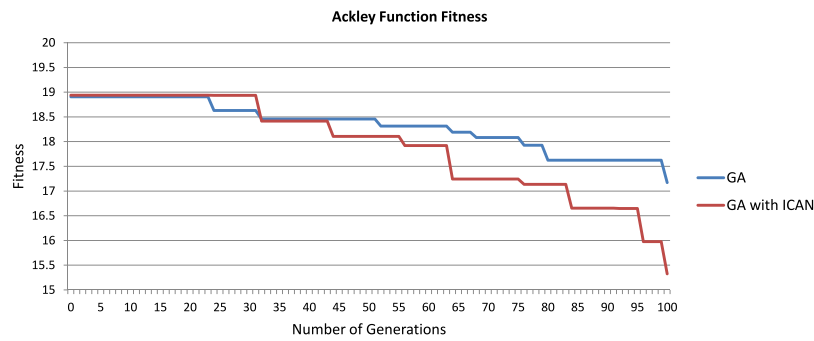
Fig. 25. Comparison between GA (existing) and GA with ICAN (proposed) approaches-based optimization using fitness of Ackley function.
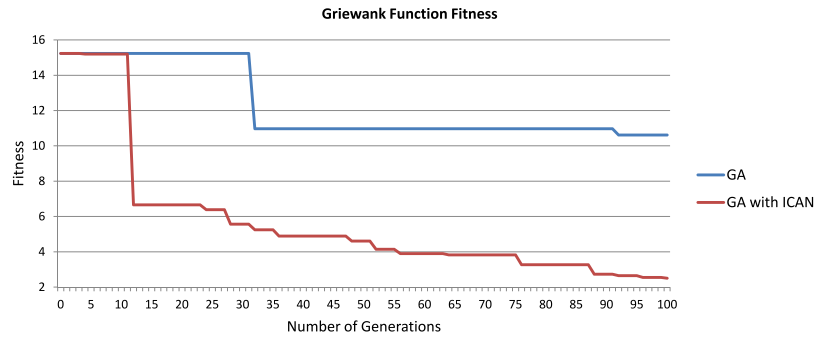

Fig. 26. Comparison between GA (existing) and GA with ICAN (proposed) approaches-based optimization using fitness of Griewank function.
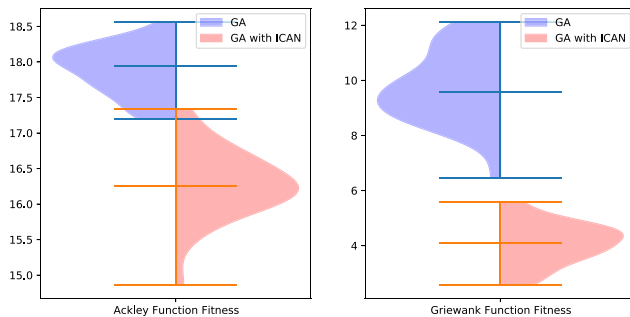

Fig. 27. Resultant data distribution in fitness of Ackley and Griewank functions for GA (existing) and GA with ICAN (proposed) approaches-based optimization.

5. Concluding remarks

In this article, we consider the inseparable concept of genetic diversity and survival of the fittest. We adopt the diversification aspect from nature and propose to integrate the same with a computational approach using GA. The proposed GA with ICAN is an extension of GA to demonstrate the significance of information fusion within metaheuristics and its application to stock price trend prediction.

As provided in the existing approaches [16,17], we append two crucial problems of parameter optimization and feature selection within a single chromosome representation of GA; though GA operations on these problems are applied individually, i.e., on part I and part II, respectively, our proposed approach enhances the searchability of GA by adaptively changing the mutation rates. ICAN also brings a conditional introduction to intra-chromosome crossover operation such that a controlled diversity is applied as per the conditions and information fusion is performed through the following inter-chromosome crossover operation. Using this approach, we optimize the window size and the number of neurons in the hidden layer(s) of LSTM prediction model through one part of chromosomes; the other part targets to select a set of features from a given list of input features [16,17]. The results indicate considerably lower performance errors using MSE, MAE, and MAPE metrics which, in turn, presents higher stock price prediction accuracy of our proposed approach. We also evaluate the stock trend prediction performance using R^2 score metric and further analyze the impact of the optimized parameter values; it is observed that with comparatively smaller window size and a significant number of neurons in LSTM hidden layers, our proposed GA with ICAN approach enhances the network's ability to predict stock price and trend. The results of the proposed GA with ICAN-based optimization approach are better than the existing GA [16, 17] as well as ALO-based optimization methods. Detailed analysis on the same demonstrates the importance of selecting a fewer number of relevant features; it can be identified that instead of 10 features as given by the reference articles [16,17], a nearly half or even lesser number of features are capable of providing essential market information and hence, reducing the cost of forecasting. Data distribution of resultant values is also analyzed using violin plots for MSE, MAE, MAPE, and R^2 score metrics that determine performance enhancement using GA with ICAN. Thus, it can be concluded that information fusion can be an essential part of GA operations and the proposed diversity-oriented GA with ICAN approach has potential extensibility to address a variety of stock market applications; the same is statistically supported with p-value less than .05 using Wilcoxon signed-rank test which, in turn, denotes adaptability of the proposed work for other application domains. We adopt Ackley and Griewank multimodal fitness functions to examine the optimization abilities of our proposed approach; a comparative study denotes remarkable convergence using GA with ICAN optimization, and hence, potential usabilities of the proposed information fusion-based optimization approach.

The proposed approach is an integration of information fusion within GA; the same can be further extended to other diversity-preserving approaches. In this article, we illustrate the application of GA with ICAN to optimize parameters and select features

from a list of input features; other feature selection, as well as construction, methods such as using information retrieval-based technique [47] can also be assessed using the proposed approach. The same can be further evaluated to check its suitability to other such optimization problems with a larger number of hyperparameter tuning as well as feature selection problems. Apart from the considered LSTM prediction model, other machine learning, as well as deep learning models, can be potentially tuned using the proposed approach; various such models have been reviewed and analyzed in some of the recent articles [48,49]. Our proposed work has been limited to the defined set of parameters and model architecture specifications for a fair comparison with the considered existing approaches [16,17], however, the prediction model can be replaced with other surrogate models that can offer fast modeling. The potential expansion of our work can be investigated with models such as multi-variate adaptive regression splines (MARS) and M5Tree for faster forecasting. While we address stock price trend prediction, other financial, as well as optimization problems, can be significantly addressed using the same. In the evaluation of the usefulness of GA [50], it can be concluded that our proposed approach can be potentially extended to other research domains including but not limited to intrusion detection system [19], human action recognition [51,52], land use optimization [53], etc.

CRediT authorship contribution statement

Ankit Thakkar: Conceptualization of the idea, Flow of the paper, Methodology, Validation of experimental results, Analysis, Investigation, Visualization, Data representation, Writing – original draft, Writing – review & editing. **Kinjal Chaudhari:** Conceptualization of the idea, Flow of the paper, Methodology, Validation of experimental results, Analysis, Investigation, Visualization, Data representation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data can be downloaded from source(s) specified in the article.

References

[1] A. Thakkar, K. Chaudhari, Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions, Inf. Fusion 65 (2021) 95–107.

[2] S. Lieberson, Measuring population diversity, Am. Sociol. Rev. (1969) 850–862.

[3] D.V. Budescu, M. Budescu, How to measure diversity when you must, Psychol. Methods 17 (2) (2012) 215.

[4] D.A. Harrison, K.J. Klein, What's the difference? Diversity constructs as separation, variety, or disparity in organizations, Acad. Manag. Rev. 32 (4) (2007) 1199–1228.

[5] L. Phillips, Human Adaptation and Its Failures, Academic Press, 2013.

[6] L.B. Scheinfeldt, S.A. Tishkoff, Recent human adaptation: genomic approaches, interpretation and insights, Nature Rev. Genet. 14 (10) (2013) 692–702.

[7] Y. Field, E.A. Boyle, N. Telis, Z. Gao, K.J. Gaulton, D. Golan, L. Yengo, G. Rocheleau, P. Froguel, M.I. McCarthy, et al., Detection of human adaptation during the past 2000 years, Science 354 (6313) (2016) 760–764.

[8] S. Mirjalili, Genetic algorithm, in: Evolutionary Algorithms and Neural Networks, Springer, 2019, pp. 43–55.

[9] S. Sivanandam, S. Deepa, Genetic algorithms, in: Introduction to Genetic Algorithms, Springer, 2008, pp. 15–37.

[10] S. Rajasekaran, G.V. Pai, Neural Networks, Fuzzy Logic and Genetic Algorithm: Synthesis and Applications (with Cd), PHI Learning Pvt. Ltd., 2003.

[11] L. Haldurai, T. Madhubala, R. Rajalakshmi, A study on genetic algorithm and its applications, Int. J. Comput. Sci. Eng. 4 (10) (2016) 139.

[12] N.F. Johnson, P. Jefferies, P.M. Hui, et al., Financial Market Complexity, OUP Catalogue, Oxford University Press, 2003.

[13] K. Chaudhari, A. Thakkar, iCREST: International cross-reference to exchange-based stock trend prediction using long short-term memory, in: International Conference on Applied Soft Computing and Communication Networks, Springer, 2021, http://dx.doi.org/10.1007/978-981-33-6173-7_22.

[14] A. Thakkar, K. Chaudhari, CREST: Cross-reference to exchange-based stock trend prediction using long short-term memory, Procedia Comput. Sci. 167 (2020) 616–625.

[15] A. Thakkar, K. Chaudhari, A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization, Arch. Comput. Methods Eng. 28 (4) (2021) 2133–2164.

[16] H. Chung, K.-s. Shin, Genetic algorithm-optimized long short-term memory network for stock market prediction, Sustainability 10 (10) (2018) 3765.

[17] S. Chen, C. Zhou, Stock prediction based on genetic algorithm feature selection and long short-term memory neural network, IEEE Access (2020).

[18] K. Chaudhari, A. Thakkar, Travelling salesman problem: an empirical comparison between ACO, PSO, ABC, FA and GA, in: Emerging Research in Computing, Information, Communication and Applications, Springer, 2019, pp. 397–405.

[19] A. Thakkar, R. Lohiya, Role of swarm and evolutionary algorithms for intrusion detection system: A survey, Swarm Evol. Comput. 53 (2020) 100631.

[20] G. Booy, R. Hendriks, M. Smulders, J. Van Groenendael, B. Vosman, Genetic diversity and the survival of populations, Plant Biol. 2 (4) (2000) 379–395.

[21] D.H. Reed, R. Frankham, Correlation between fitness and genetic diversity, Conserv. Biol. 17 (1) (2003) 230–237.

[22] J. Puy, C.P. Carmona, H. Dvořáková, V. Latzel, F. de Bello, Diversity of parental environments increases phenotypic variation in Arabidopsis populations more than genetic diversity but similarly affects productivity, Ann. Botany 127 (4) (2021) 425–436.

[23] D. Gupta, S. Ghafir, An overview of methods maintaining diversity in genetic algorithms, Int. J. Emerg. Technol. Adv. Eng. 2 (5) (2012) 56–60.

[24] D.K. Sharma, H. Hota, K. Brown, R. Handa, Integration of genetic algorithm with artificial neural network for stock market forecasting, Int. J. Syst. Assur. Eng. Manag. (2021) 1–14.

[25] F. Weng, H. Zhang, C. Yang, Volatility forecasting of crude oil futures based on a genetic algorithm regularization online extreme learning machine with a forgetting factor: The role of news during the COVID-19 pandemic, Resour. Policy 73 (2021) 102148.

[26] Y. Huang, Y. Gao, Y. Gan, M. Ye, A new financial data forecasting model using genetic algorithm and long short-term memory network, Neurocomputing 425 (2021) 207–218.

[27] F. Ghashami, K. Kamyar, et al., Performance evaluation of ANFIS and GA-ANFIS for predicting stock market indices, Int. J. Econ. Finance 13 (7) (2021) 1.

[28] A. Javadian Kootanaee, A.A. Poor Aghajan, M. Hosseini Shirvani, A hybrid model based on machine learning and genetic algorithm for detecting fraud in financial statements, J. Optim. Ind. Eng. 14 (2) (2021) 169–186.

[29] S.H. Jeong, H.S. Lee, H. Nam, K.J. Oh, Using a genetic algorithm to build a volume weighted average price model in a stock market, Sustainability 13 (3) (2021) 1011.

[30] Z.J. Ye, B.W. Schuller, Capturing dynamics of post-earnings-announcement drift using a genetic algorithm-optimized xgboost, Expert Syst. Appl. 177 (2021) 114892.

[31] M.-S. Cheong, M.-C. Wu, S.-H. Huang, Interpretable stock anomaly detection based on spatio-temporal relation networks with genetic algorithm, IEEE Access 9 (2021) 68302–68319.

[32] E.K.W. Leow, B.P. Nguyen, M.C.H. Chua, Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation, Expert Syst. Appl. 179 (2021) 115060.

[33] I. Yaman, T.E. Dalkılıç, A hybrid approach to cardinality constraint portfolio selection problem based on nonlinear neural network and genetic algorithm, Expert Syst. Appl. 169 (2021) 114517.

[34] S.K. Mittal, N. Srivastava, Mean-variance-skewness portfolio optimization under uncertain environment using improved genetic algorithm, Artif. Intell. Rev. 54 (8) (2021) 6011–6032.

[35] C.-H. Chen, Y.-H. Chen, V.G. Diaz, J.C.-W. Lin, An intelligent trading mechanism based on the group trading strategy portfolio to reduce massive loss by the grouping genetic algorithm, Electron. Commer. Res. (2021) 1–40.

[36] N. Adhikari, M. Sultana, S. Bhattacharya, An approach of safe stock prediction using genetic algorithm, in: Proceedings of International Conference on Innovations in Software Architecture and Computational Systems, Springer, 2021, pp. 119–138.

[37] Y. He, H. Li, W. Wei, Application of LSTM model optimized based on adaptive genetic algorithm in stock forecasting, in: International Conference on Data Mining and Big Data, Springer, 2021, pp. 46–58.

[38] Yahoo Finance, Yahoo finance, 2022, https://in.finance.yahoo.com. (Last Accessed 14 March 2022).

[39] Yahoo Finance, Yahoo finance, 2021, https://in.finance.yahoo.com/quote/%5EKS11/history?p=%5EKS11. (Last Accessed 02 July 2021).

[40] H. Zhang, H. Lin, Y. Li, Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification, IEEE Geosci. Remote Sens. Lett. 12 (5) (2015) 1061–1065.

[41] D. Thakor, A. Shah, D_EDF: An efficient scheduling algorithm for real-time multiprocessor system, in: 2011 World Congress on Information and Communication Technologies, IEEE, 2011, pp. 1044–1049.

[42] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, An innovative neural network approach for stock market prediction, J. Supercomput. (2018) 1–21.

[43] B.H. Abed-Alguni, D.J. Paul, Hybridizing the cuckoo search algorithm with different mutation operators for numerical optimization problems, J. Intell. Syst. 29 (1) (2020) 1043–1062.

[44] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications, Future Gener. Comput. Syst. 97 (2019) 849–872.

[45] H. Emami, Stock exchange trading optimization algorithm: a human-inspired method for global optimization, J. Supercomput. 78 (2) (2022) 2125–2174.

[46] M. Locatelli, A note on the Griewank test function, J. Global Optim. 25 (2) (2003) 169–174.

[47] A. Thakkar, K. Chaudhari, Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks, Appl. Soft Comput. 96 (2020) 106684.

[48] A. Thakkar, D. Patel, P. Shah, Pearson Correlation Coefficient-based performance enhancement of Vanilla Neural Network for stock trend prediction, Neural Comput. Appl. (2021) 1–16.

[49] A. Thakkar, K. Chaudhari, A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions, Expert Syst. Appl. (2021) 114800.

[50] S. Katoch, S.S. Chauhan, V. Kumar, A review on genetic algorithm: past, present, and future, Multimedia Tools Appl. 80 (5) (2021) 8091–8126.

[51] P. Pareek, A. Thakkar, RGB-D based human action recognition using evolutionary self-adaptive extreme learning machine with knowledge-based control parameters, J. Ambient Intell. Humaniz. Comput. (2021) 1–19.

[52] P. Pareek, A. Thakkar, An improved method for classifying depth-based human actions using self-adaptive evolutionary technique, J. Ambient Intell. Humaniz. Comput. (2022) 1–17.

[53] X. Ding, M. Zheng, X. Zheng, The application of genetic algorithm in land use optimization research: A review, Land 10 (5) (2021) 526.