A Novel Framework for Drug Synergy Prediction using Differential Evolution based Multinomial Random Forest

Jaspreet Kaur[1], Dilbag Singh[2], Manjit Kaur[3]

Computer Science & Engineering,
Apex Institute of Technology
Chandigarh University
Gharuan, Punjab, India

Abstract—An efficient prediction of drug synergy plays a significant role in the medical domain. Examination of different drug-drug interaction can be achieved by considering the drug synergy score. With an rapid increase in cancer disease, it becomes difficult for doctors to predict significant amount of drug synergy. Because each cancer patient's infection level varies. Therefore, less or more amount of drug may harm these patients. Machine learning techniques are extensively used to estimate drug synergy score. However, machine learning based drug synergy prediction approaches suffer from the parameter tuning problem. To overcome this issue, in this paper, an efficient Differential evolution based multinomial random forest (DERF) is designed and implemented. Extensive experiments by considering the existing and the proposed DERF based machine learning models. The comparative analysis of DERF reveals that it outperforms existing techniques in terms of coefficient of determination, root mean squared error and accuracy.

Keywords—Machine learning; random forest; drug synergy

I. INTRODUCTION

With changing lifestyle, more and destructive diseases are occurring because of poor dietary habits, absence of physical exercises, alcohol utilization, etc. Most common harmful diseases nowadays are cancer, obesity, heart disease, stroke and type II diabetes. Utilization of single drug gives just single supplement; however experiencing big illnesses likewise influences entire body [1]. Therefore, combination of different drugs is required for giving appropriate treatment and legitimate supplements to the body. Mix of various drugs is essentially relies on two things i.e., drug oriented or disease oriented. If mixture is drug oriented then it will concentrate on combination of drugs without knowing disease [1]. However, if it is disease oriented, then many different combinations of drugs are possible. In this way, making pair of various drugs and blends isn't easy. Forecast of drug synergy score is an ill caused issue [2]. For this purpose, various machine learning techniques are also implemented, many techniques are also compared in terms of different parameters like accuracy [2]. Prediction of drug synergy is very important to prevent many harmful diseases. It assumes an effective job in medical domain for preventing particular cancer agents. Machine learning methods has a capacity to limit the synergy estimation errors and thus, used ensemble based differential evolution to optimize the SVM regression technique. Developed a Synergistic field-aware factorization machine SyFFM which uses pharmacological data

to inspect and forecast different combinations of drugs [3]. While making different drug combinations, it is very important to contemplate various sources of information like chemical, biological, pharmacological and network knowledge [4]. To minimize the side effects of different combinations of drugs, it is very important to contemplate natural belongings and network knowledge of drugs which helps to find efficient drug combinations [4]. Combining various drugs also leads to drug toxicity and analyzed drug combinations over in vitro normal cell lines. Also, produced various combinations whose effect on normal cell lines is less [5]. Multiple drug combination treatment is much more effective than single drug. Therefore, various Ayurveda complex combinations are explored with their benefits [6]. Also, features made of DNN which are of high-level are more powerful than carefully assembled features for predicting the cell penetrability of fundamentally assorted synthetic mixes in Caco-2 cell lines [7]. Current vaccine adjuvants featuring and the benefit of immune drug synergy to adjuvant and immunization model is made. The attention was on new advances which are studied and applied adjuvant on immune synergies and immunization improvement [8]. Taiji has been exhibited, software with a high-performance for quick and precise estimation of drug synergy dependent on the winning algorithm [9]. Also, synergy prediction work process can carry compound prioritization in huge scale medicate screenings, and synergy stratification work process can choose where the viability of medications definitely known for inciting synergy is higher [10].

A computational framework biological device is built, DrugComboExplorer, to distinguish pathways of driver signaling and anticipate combinations of synergistic drug by incorporating the learning inserted in huge measures of accessible omics and pharmacogenomics information [11]. Drug synergy, of numerous types, can be anticipated with high degrees of exactness with significant clinical potential [12]. This leap understanding of joint systems of action will take into account the plan of balanced combinatorial therapeutics on a vast scale, across various cancer types. A synergy score dependent on the contrast between the drug and the single drug dose response- bends. The CSS-based synergy score can distinguish genuine synergistic and antagonistic drug combinations. An exploratory computational pipeline has been depicted, named as target addition scoring (TAS), that numerically converts the profiles of drug response to target fixation marks, and

in this way gives a ranking of potential therapeutic targets as indicated by their utilitarian significance in a specific cancer sample [13]. A complete survey of the different drug-re purposing techniques concentrating on the computational methodology [14]. Mechanism synergy estimation used well-distinguished information of biology to anticipate interaction of drugs dependent on medication target interactions [15].

Contributions: Following are our main contributions in this paper:-

1) Initially, various machine learning techniques are analyzed to estimate drug synergy score.
2) Thereafter, to overcome the issue or parameters tuning issue, in this paper, an efficient Differential evolution based multinomial random forest (DERF) is designed and implemented.
3) Comparative analysis are also drawn between the existing and the proposed machine learning models in terms of coefficient of determination, root mean squared error and accuracy.

The rest of paper is summarized as follows: Section II discusses Related work. Mathematical preliminaries are given in Section III. Section IV provides description of proposed technique. Section V demonstrates Performance analysis. Conclusion remarks are discussed in Section VI.

II. RELATED WORK

In this section, Related work has been demonstrated which is as follows: For the analysis of drug combinations Sałat et al., 2015, proposed an universal support vector regression (SVR)-based technique that significantly increases the isobolographic investigation [16]. Until drugs are endorsed, many side effects are not perceivable in clinical preliminaries . Therefore, Zhang et al., 2016, formulated drugs that are endorsed, drug and terms of side effect– symptom relationship as a recommend-er framework, and change the issue of estimating side effects into a recommend-er task [17]. Weinstein et al., 2017, Substantial number of drug combinations are given, to organize the best treatments with computational strategies are critical and presented different methodologies to predict synergistic drug interactions [15]. Wang et al., 2017, proposed another computational technique from drug molecular structure for estimating DTIs and protein succession by utilizing the stacked auto-encoder of deep learning which can sufficiently separates the crude information data. Conventional machine learning methods to deal with sensitivity of drug estimation accept that training information and test information must be in a similar feature space and have the equivalent basic distribution [18]. Two methods of exchange learning are presented by Turki et al., 2017, which combine the auxilliary information from the related undertaking with the training information of the objective task to enhance the forecast execution of the target task on the test information [19]. Tsigelny et al., 2018, examined a few machine learning techniques that have been effectively executed in a few instances of combination drug therapy from hypertension, HIV, irresistible sicknesses to cancer [20]. Hemalatha et al., 2018, considered a survey of methodologies that have been presented to handle sensitivity of drug estimation issue particularly as for the customized Cancer treatment

[21]. Yuan et al., 2018, aimed to build up an applicable new SVM model by consolidating the majority of the highlights of the atomic property-based descriptors and fingerprints to improve the precision for the BBB permeability prediction [22]. Olier et al., 2018, examined the learning of quantitative structure activity relationships (QSARs) as a contextual analysis of meta-learning [23]. This area of application is the most noteworthy societal significance, because it is main step to develop new medicines. Levine et al., 2018, examined how slacked linear regression can be utilized to distinguish the physiologic impacts of drugs from information in the electronic health record (EHR) [24]. Tom et al., 2018, concentrated on new advances and applied immune synergies to adjuvant and vaccine improvement. Su et al., 2019, presented Deep-Resp-Forest that has exhibited the promising utilization of deep learning and deep forest approach on the drug reaction estimation tasks [25]. As per Li et al., 2019, Spectroscopy of near-infrared joined with chemometrics was utilized to analyze the fundamental dynamic components including caffeic acid, chlorogenic acid, ursodesoxycholic acid, luteoloside, chenodeoxycholic acid and baicalin in the Tanreqing injection [26]. Mofrad et al., 2019, created and approved a clinically pertinent decision tree in the finding of Alzheimer's disease (AD) for the utilization of cerebrospinal liquid biomarkers [27]. Strategy is proposed by Bashar et al., 2019, to assess the Heart rate (HR) from wearable gadgets using random forest algorithm [28]. Ogunleye et al., 2019, given another information partitioning rule utilizing the mean of the information sections to develop the tree until the child nodes are little in size [29]. Then, connection is made between the local regression and leave nodes to improve the goals of the node outputs. Randomization is presented at tree development and creation of forest. Zhao et al., 2019, created and approved a estimation model by information extracted of eGFR from a territorial health framework [30]. Feng et al., 2019, proposed a strategy for support vector machine dependent on the dragonfly algorithm (DA-SVM) in a offshore oil field to estimate the short-team load of the microgrid [31]. As per Kestenbaum et al., 2019, Regression is a numerical method used to evaluate the relationship between two or more study variables [32]. Norman et al., 2019, Reported flu immunization inclusion in kids with medical comorbidities remains inadequate. Through the investigation and examination of the information of the department of respiration that is combined with the information of the quality of air measurement, meteorological measurement, and time measurement, Jin et al., 2019, considered its related features and sets up a multidimensional estimation features model dependent on a BP neural system [33]. An attempt has been made by Rajalakshmy et al., 2019 to extract a couple of important time area features from sEMG signals. [34]. The examination by Hazelden et al., 2019, identified a few factors which might be helpful to recognize patients as high risk for hospitalization and the following stages will be to decide and consider the role of the drug specialist in preventing hospitalization of these patients [35]. However, machine learning based drug synergy prediction approaches suffer from the parameter tuning problem. Also, these techniques suffer from poor computational speed.

III. Mathematical Preliminaries

In this section, Multinomial random forest (MRF) has been presented. Also, Multinomial random forest is compared with random forest.

A. Random Forest for Regression

There are three aspects in which random forest and Multinomial random forest (MRF) is different. In place of bootstrap technique, partitioning procedure of training set has been used. In attribute set choice, randomness has been introduced at each internal node, and for selecting splitting point , an impurity-based multinomial distribution has been utilized. By doing these, we guarantee that each attribute and each conceivable splitting point get an opportunity for the selection. The key is to limit the negative effect of the randomness, which is required for consistency, on prediction execution.

Suppose $\mathcal{Z}_n$ indicates a data set consisting of n $r.r.d.$ examinations. Every examination has the form (P,Q), in which $P \in R^Z$ denotes the Z-dimensional attributes and $Q \in \{1, \cdots, N\}$ is the correlating mark of the examination.

1) Training Data Set Partition: To make a tree, the training set is isolated arbitrarily into two non-overlapping subsets. Different roles have been played by the two subsets. To make the structure of a tree, one subset has been used; the samples in this subset are called as structure points. Once a tree is made, on the premises of the other subset, the marks of its leaves will be re-directed; the samples in this subset are called as estimation points. In this procedure, the structure points are utilized just to change the shape of the tree by deciding the splitting point in each interior node, and for the final prediction, the estimation points have been utilized. For guaranteeing consistency of the tree, the training set's partitioning and the detachment of their roles are essential.

To assemble another tree, the training set is re-divided arbitrarily and freely. The sizes of the two subsets are fixed. The proportion of the two sections is parameterized by partition rate = $|\text{Structure points}|/|\text{Estimation points}|$.

2) Tree Construction: When we compare the proposed Multinomial random forest (MRF) with the original Random forest (RF) then in place of the bootstrap technique, partitioning of the training set has been done. In a rational way, some randomness has been introduced while selecting candidate features and splitting point. There are different ways to select the candidate features, but each feature has to be selected with positive likelihood at every split. The case has been discussed with the use of Bernoulli distribution for more appropriate comparison.

The initial change in Multinomial random forest (MRF) is to randomly select candidate features. Draw randomly from a Bernoulli distribution $C(t)$ especially for every interior node. Randomly select it as a feature set if it is 1; else move with the native process for feature set selection (*e.g.* choose randomly $O(Z)$ features, where $O(Z) = \sqrt{Z}$ or $O(Z) = \log Z$).

In the problem of classification , decrease in the impurity is caused by splitting point v is represented by:

$$V(w) = L(\mathcal{Z}^K) - \frac{|\mathcal{Z}^{K_l}|}{|\mathcal{Z}^K|}L(\mathcal{Z}^{K_l}) - \frac{|\mathcal{Z}^{K_r}|}{|\mathcal{Z}^K|}L(\mathcal{Z}^{K_r}), \quad (1)$$

where $\mathcal{Z}^K$ is the structure points and $\mathcal{Z}^{K_l}, \mathcal{Z}^{K_r}$ are two children sets created by $\mathcal{Z}^K$ splitting at w, $L(\cdot)$ is the criterion of impurity (*e.g.* Shannon entropy or Gini index).

The other change is that in-spite of the deterministic rule, there is random selection of the splitting point on the basis of a multinomial distribution. In original random forest maximization of $V(w)$ used to be done where splitting point w is intended. But here, Splitting points are selected randomly as per multinomial distribution $Q(\phi)$ made on the basis of the impurity decrease of all possible points. The certain setting up of $Q(\phi)$ is given below:

For all possible splitting points, suppose $V = (V_1, \cdots, V_m)$ be the vector of impurity decline and all candidate features. Firstly, normalize it as $\hat{V} = (\frac{V_1 - \min V}{\max V - \min V}, \cdots, \frac{V_m - \min V}{\max V - \min V})$, and then compute the probabilities $\alpha = (\alpha_1, \cdots, \alpha_m) = \text{softmax}(C\hat{V})$, where $C > 0$ is a hyper parameter.

For selecting the splitting point, to regulate the probabilities the hyper parameter C plays an important role. If C is larger, then there will be more probability to select largest impurity decrease point. The MRF totally becomes random forest when $t \to 0$ and $C = 0$, for the splitting point selection process. The MRF turns Breiman's random forests when $t \to 0$ and $C \to +\infty$.

From the above two processes, To grow a tree, Selection of one feature and its correlating splitting value is done. Structure points influence only the building of the tree while for prediction estimation points are being involved. The process of splitting is to be continued until the given halting criteria are fulfilled.

Just like random forests, MRF's halting condition is also associates to the minimum leaf size. More specifically, in every leaf, the number of estimation points is required to be larger than g_e where $g_e \to \infty$ and $g_e/e \to 0$ as $e \to \infty$.

3) Prediction: When a tree has been developed utilizing the structure points as portrayed above, we re-decide the predicted values for the leaves utilizing the estimation points.

Suppose the unlabeled sample is j and f denotes a decision tree in MRF. The probability that the sample j with label $d(d \in \{1, \cdots, N\})$ evaluated by this tree is

$$\beta^{(d)}(j) = \frac{1}{|\mathcal{O}_f^H(j)|} \sum_{(P,Q) \in \mathcal{O}_f^H(j)} Q = d, \quad (2)$$

and the prediction is given by maximizing $\beta^{(d)}(j)$:

$$\hat{b} = f(j) = \arg \max_d \{\beta^{(d)}(j)\}, \quad (3)$$

where $\mathcal{N}_f^H(j)$ represents the number of estimation points in the node containing j, $(\cdot)$ is the indicator function.

The last prediction from the MRF depends on the greater part vote:

$$\bar{\hat{b}} = \overline{f^{(Q)}(j)} = \arg \max_d \sum_{i=1}^{Q} f^{(i)}(j) = d, \quad (4)$$

where Q is the number of separate trees in Multinomial random forest(MRF).

EDE uses adaptive $U \in [0.1, 1.0]$ and $CR \in [0, 1]$ On the basis of their last execution, the mutation method and the features of CR are self-organized. Mutation is decays in two types depending upon the proposed technique. Best population is given by first one but the other does not have. In random fashion from these types, selection of two methods by proposed technique is done during evolution. Hence, the ensemble based DE performs better from variants of DE which are already existing.

B. Standard Differential Evolution

One of the meta-heuristic technique is Differential evolution (DE), which is a easy, efficient and powerful global optimization method. When compared with the competitive optimization methods in case of convergence speed along with robustness, DE performed better in many real world applications.

To optimize a given problem, DE uses recombination, mutation and selection. In the beginning, population is created randomly. To create new solutions, recombination, mutation and selection operators are used. For the evaluation of an optimistic trial vector, the selection operator is used for next iteration.

DE begins with a population of e_t D-dimensional candidate solutions, that are presented as:

$$\Pi_{k,\delta}(k = 1, 2, \ldots, e_t) = n_{k,\delta}^1, n^2 k, \delta, \ldots, n_{k,\delta}^Z \quad (5)$$

where k denotes the population with the kth solution. δ presents current generation. ω denotes the population dimension.

Initially, population focuses to consider every search domain and bounded as:

$$\Pi_{lb} = n_{lb}^1, n_{lb}^2, \ldots, n_{lb}^\omega \quad (6)$$

Also, population in the beginning is constrained to use maximum bound. Maximun bound is presented as:

$$\Pi_{ub} = n_{ub}^1, n_{ub}^2, \ldots, n_{ub}^\omega \quad (7)$$

Thus, initially, population can be rewritten as:

$$n_{k,0} = m_{lb} + r1 * (n_{ub} - n_{lb}) \quad (8)$$

Here, uniformly distributed random variable is represented as r1. where r1=$rand(0,1) \in [0,1]$.

1) Mutation: In the second step, Mutation is utilized to evaluate a trail vector as:

$$W_{k,\delta} = w_{k,\delta}^1, w_{k,\delta}^2, \ldots, w_{k,\delta}^\omega \quad (9)$$

Using different mutation methods, the $W_{k,\delta}$ is evaluated:

Q_c^1:
$$W_{k,\delta} = \Pi_{\text{best},\delta} + G.(\Pi_{h_1^k,\delta} - \Pi_{h_2^k,\delta}) \quad (10)$$

Q_c^2:
$$W_{k,\delta} = \Pi_{\text{best},\delta} + G.(\Pi_{h_1^k,\delta} - \Pi_{h_2^k,\delta}) + G.(\Pi_{h_3^k,\delta} - \Pi_{h_4^k,\delta}) \quad (11)$$

Q_c^3:
$$W_{k,\delta} = \Pi_{k,\delta} + G.(\Pi_{\text{best},\delta} - \Pi_{k,\delta}) + G.(\Pi_{h_1^k,\delta} - \Pi_{h_2^k,\delta}) \quad (12)$$

Q_c^1:
$$W_{k,\delta} = \Pi_{h_1^k,\delta} + G.(\Pi_{h_2^k,\delta} - \Pi_{h_3^k,\delta}) \quad (13)$$

Q_c^2:
$$W_{k,\delta} = \Pi_{h_1^k,\delta} + G.(\Pi_{h_2^k,\delta} - \Pi_{h_3^k,\delta}) + G.(\Pi_{h_4^k,\delta} - \Pi_{h_5^k,\delta}) \quad (14)$$

Q_c^3 :
$$v_{k,\delta} = \Pi_{k,\delta} + N.(\Pi_{h_1^k,\delta} - \Pi_{h_k,\delta}) + G.(\Pi_{h_2^k,\delta} - \Pi_{h_3^k,\delta}) \quad (15)$$

Here, $h_1^k, h_2^k, h_3^k, h_4^k, h_5^k \in [0, 1]$ represent mutually exclusive indexes. To constraint the amplification of DE, G presents mutation scale factor.

2) Recombination: To maximize the target vectors diversity, the operator for recombination is used. Implementation of recombination process is given below:

$$W_{k,\delta} = w_{k,\delta}^1, w_{k,\delta}^2, \ldots, w_{k,\delta}^\omega$$
$$\Pi_{k,\delta} = \Pi_{k,\delta}^1, \Pi_{k,\delta}^2, \ldots, \Pi_{k,\delta}^\omega \quad (16)$$

To make a trial vector, Eq. 16 is utilized as:

$$v_{k,\delta} = v_{k,\delta}^1, v_{k,\delta}^2, \ldots, v_{k,\delta}^\omega \quad (17)$$

Thus, $(v_{k,\delta}^k)$ is computed as a new trial vector:

$$v_{k,\delta}^k = \begin{cases} w_{k,\delta}^k & \text{if } \text{rand}_k[0,1) \leq \text{CR or } (k = k_{\text{rand}}) \\ n_{k,\delta}^k & \text{others} \end{cases} \quad (18)$$

Here, the recombination constant is presented as $CR \in [0, 1]$. A randomly selected index is $jrand \in [1, \omega]$ which makes sure that $v_{k,\delta})$ will be different from $\Pi_{k,\delta}$ by not less than one parameter.

3) Selection: If $v_{k,\delta}^k$ is more than the upper or lower limits, then within the search range, random re-initialization will be done. Then, the fitness values of all trail vectors $(v_{k,\delta}^k)$ are evaluated as:

$$n_{k,\delta+1} = \begin{cases} v_{k,\delta} & \text{if } f(v_{k,\delta}) \leq f(n_{k,\delta}) \\ n_{k,\delta} & \text{otherwise} \end{cases} \quad (19)$$

In case, if better fitness is given by $v_{k,\delta}$ when compared to $n_{k,\delta}$, then replacement of $v_{k,\delta}$ will be done with $n_{k,\delta}$ and further proceed for succeeding generation; else continue .

IV. PROPOSED TECHNIQUE

In this section, Proposed technique has been explained in detail. Algorithm of Proposed technique also been demonstrated .

A. Ensemble Mutation Operator

Performance of DE based upon the above operators, which are recombination and mutation. Further, mutation operators are separated into two types (i.e, Q_h^g and Q_c^g). The Q_c^g is with best solution in in Eqs. (5-7) , and the Q_h^g is without best solution including Eqs. (8-10). For the diversity of the population to be balanced and to balance the convergence speed, two mutation operators are considered by the proposed technique. The first one is obtained from Q_h^g, and second one is obtained from Q_c^g. Two mutation operators are randomly selected from Q_h^g and Q_c^g, respectively.

Algorithm 1: EDE based synergy prediction Step 1:

1) First of all, parameters initialization of proposed ensemble based DE is done (i.e, population size (e_t), function evaluations (ρ), two mutation operators obtained from Eqs. (10-12) and Eqs. (13-15), respectively. Also, maximum number of ρ are represented by Q_ρ

2) Initialize a counter $\delta = 0$. Randomly initialize a population with size e_t (i.e, $\nu = \theta_{1,\delta}, \ldots, \theta_{e_t,\delta}$) with $\theta_{k,\delta} = n_{k,\delta}^1, \ldots, n_{k,\delta}^\lambda, k = 1, \ldots, e_t$ uniformly distributed within $[\theta_{lb}, \theta_{ub}]$, where $\theta_{lb} = n_{lb}^1, n_{lb}^2, \ldots, n_{lb}^\lambda$ and $\theta_{ub} = n_{ub}^1, n_{ub}^2, \ldots, n_{ub}^\lambda$.

3) Evaluate the fitness of each population and determine the solution with best fitness (b_s).

4) $\rho = \rho + e_t$;

5) While $\rho \leq Q_\rho$
   for $k = 1 : e_t$
   ```
           Compute W_{k_1,δ} by the first
   technique
           Compute W_{k_2,δ} by the second
   technique
           Compute trial vector v_{k_1,δ} by
   Eq.(19)
           Compute trial vector v_{k_2,δ} by
   Eq.(19)
             ρ = ρ+2
   End for
   ```

6) If any variable is outside its limits, then re-initialization of trial vector $v'_{k,\delta}$ (including $v_{k\_1,\delta}$ and $v_{k\_2,\delta}$) within the search space is done randomly.

7) Selection Procedure
   for j=1 to e_t
   ```
           Evaluate the trial vector v_{k,δ}
            If  w(v'_{k,δ}) ≤ w(θ_{k,δ})
            θ_{k,δ+1} = v'_{k,δ}, w(θ_{k,δ+1}) = w(v'_{k,δ})
            P_{k,tbest} = v'_{k,δ}, w(P_{k,tbest}) = w(v'_{k,δ})
             If  f(v'_{k,δ}) < w(θ_{best,δ})
             θ_{best,δ} = v'_{k,δ}, w(θ_{best,δ}) = w(v'_{k,δ})
             t_g = v'_{k,δ}, f(t_g) = w(v'_{k,δ})
            End if
           End if
   End for
   ```

8) The generation count $\delta = \delta + 1$ is Incremented;

9) End while

In this paper, Accuracy(ACC) and correlation coefficient(CCO) are utilized to evaluate the fitness of every solution as given below:

$$Maximize(ACC, CCO) \tag{20}$$

B. Synergy Prediction using EDE

Various steps are portrayed that are needed to be implement to assess the synergy prediction in a productive way. Different steps are given below:

1) **Selection of framework :** For the completion of the presented procedure, multinomial random forest framework is used. Following parameters will be optimized with the use of proposed technique.

2) **Scaling:** It is used to ignore characteristics which are in highest numeric limit from the minimum numeric limit. It likewise decreases the computational complexity for machine learning methods. scaling of numeric attributes between either $[-1, 1]$ or $[0, 1]$ is done as follows in this paper:

$$w' = \frac{w - l_c^b}{u_c^b - l_c^b} \tag{21}$$

Here, w represents native value. w' is a scaled value. u_b and l_b represent upper and lower limit of feature values, respectively.

3) **Training and testing data:** The synergistic information is deteriorated into training and testing information. To prepare the proposed random forest based machine learning technique, the training information is used. After that, For the proposed method, testing information is utilized to screen the viability.

4) **Differential evolution based RF:**The differential evolution based random forest is utilized for the tuning of the required attributes of random forests. As a fitness function, the Root mean squared blunder (RMSE) is used. The general target is to discover best parameters for random forest based synergy prediction strategy.

5) **Execution criteria:** At the point when the Execution criteria accomplished, the differential evolution based RF ends itself and then return tuned characteristics; else, for other iterations of DE continues .

6) **Method building:** The end results acquired from DE based random forest are utilized as Random forest technique's initial attributes. Then, it is utilized to make the trained Random forest based synergy estimation method.

7) **Performance analysis:** Here, trained method acquired utilizing DE based random forest. Random forest is put in on the training information to foresee the drug synergy esteems. Subsequently, Use these values for assessing the execution of presented machine learning method.

V. PERFORMANCE ANALYSIS

This section gives the comparative analysis of proposed and existing machine learning strategies.The data of drug synergy score, which has been used for validation purpose in this research work, comprise of two different terms as Highest concentration of drug P and Highest concentration of drug Q. On distinct drug's distinct concentrations, the benefit of

drug synergy data set relies. Thus, Implemented the proposed model on the train dataset which comprises of different drugs in different concentration for drug synergy prediction.The different drugs in different concentrations has been assessed using Drug interaction coefficient (DIC), which is assessed as follows:

$$DIC = \frac{PQ}{P \times Q} \qquad (22)$$

Here, PQ is the extent to the control combination from the two-drug mix combination and P or Q is the extent to the control combination from the single drug combination . The $DIC < 1$ shows synergism, particularly $DIC < 0.7$ demonstrates a fundamentally synergistic impact, $DIC = 1$ demonstrates additivity and DIC > 1 shows opposition.

In this paper, 15-overlap cross-approval is utilized for the testing of ensemble based machine learning exhibit at the time of preparation stage to assess over fitting issue. To accomplish, 15-overlap cross-approval, at first training information has been partitioned into 15 equal subsets (overlap). Approval set is used to keep the 1-overlap and keep other 14-overlaps in the cross approval preparing set. Demonstrated by utilizing the cross-approval set of training and assess the precision of the presented model by favoring the estimated values which opposes the endorsement set. Correspondingly, precision of each of the 15-overlaps have been assessed. To conquer the issue of over-fitting, mean of assessed ac-curacies have been assessed. For approval, each overlap has been utilized just once. In this way, 15-overlay cross-approval ensures, the presented trained model does not experiences the over-fitting problem. The proposed DERF and the existing methods are implemented in Python 3.6. Intel core i5 8th generation processor is used along with 8GB RAM and 2GB graphics card. Here, 20% to 90% ratio of air pollution dataset is taken for training purpose. Also, rest of dataset is used for testing purpose. The acceptance error is allowed only between , for accuracy evaluation.

A. Experimental Setup

DERF with another methods have been executed with 16 GB RAM on Intel core $i7$ processor. For designing simulation environment, software, MATLAB $2017a$ is used with matplotlib library. In next section, datasets description(i.e., AstraZeneca-Sanger Drug Combination) [36] with its characteristics is given.

B. Dataset

Various drug combinations are performed to evaluate the drug impacts on cells at various concentrations. The concentration space has been increased in the presence of two drugs: by dimension and the induced effect which has been described by a dose reaction surface that opposes to a curve. A dose reaction surface will generally resemble this: The synergy score information, which has been utilized for approval reason contain 2 terms for example (1) Highest concentration of drug P and (2) Highest concentration of drug Q. Thus, the benefit of synergy data is based on the different drugs combinations. DERF has been actualized on the train data which contains

the distinctive grouping of various medications to analyze the score of synergy . Various drugs with different concentrations has been assessed by utilizing drug interaction coefficient (DIC). Various features of drug synergy dataset are described in Table I.

TABLE I. CHARACTERISTICS OF DRUG SYNERGY DATASET

Column name	Explanation
COMP_P	Drug P's name.
COMP_Q	Drug Q's name.
COMB_ID	Name of the combination of drug P and drug Q
C_L	Normalised cell line name.
IC50_P	Concentration in which half of the highest number of elimination is acquired with drug P.
H_P	Dose-reaction curve's slope for drug P.
Einf_P(Potency)	Highest number of cells killed (percentage) with drug P.
IC50_Q	Concentration in which half of the highest number of elimination is acquired with drug Q.
H_Q	Dose-reaction curve's slope for drug Q.
HIGH_CONC_P	Highest concentration of drug P.
HIGH_CONC_Q	Highest concentration of drug Q.
QA	Assurance of quality flag of combination assays
SYN_SCORE	Evaluated overall synergy of drug P and drug Q in mix.
Einf_Q	Highest number of cells killed (percentage) with drug Q.

C. Over and Under-Fitting Evaluation

In this section, Depending upon the same fraction of data, various machine learning methods are trained and tested. It is mainly used for the evaluate the over-fitting and under-fitting issue.

Tables II, III, and IV depict the performance analysis of the already existing and the proposed machine learning methods. It is found that DERF outperforms existing methods in terms of accuracy, coefficient of determination, and Root mean squared error, respectively.

TABLE II. TRAINING ACCURACY ANALYSIS

Dataset	20 %	40 %	60 %	80 %
LR	88.1 ±1.3	91.8 ±1.1	87.7 ±0.9	88.5 ±1.4
DT	89.3 ±1.9	92.4 ±1.1	88.4 ±1.8	89.8 ±2.2
RF	90.2 ±1.3	92.7 ±1.1	89.0 ±1.4	89.8 ±1.6
SVM	90.2 ±1.7	92.7 ±0.9	89.0 ±2.4	89.8 ±2.1
ANN	93.7 ±0.8	94.6 ±1.1	91.4 ±1.2	93.6 ±0.9
ANFIS	94.3 ±0.8	95.4 ±0.8	91.9 ±1.0	94.6 ±1.2
DERF	98.4 ±0.7	99.4 ±0.4	97.3 ±0.7	99.4 ±0.5

TABLE III. TRAINING COEFFICIENT OF DETERMINATION ANALYSIS

Dataset	20 %	40 %	60 %	80 %
LR	0.87 ±0.08	0.87 ±0.04	0.82 ±0.02	0.82 ±0.06
DT	0.88 ±0.01	0.87 ±0.02	0.83 ±0.06	0.83 ±0.04
RF	0.88 ±0.06	0.88 ±0.02	0.83 ±0.06	0.83 ±0.02
SVM	0.90 ±0.03	0.89 ±0.07	0.85 ±0.02	0.85 ±0.09
ANN	0.91 ±0.04	0.90 ±0.02	0.86 ±0.09	0.86 ±0.05
ANFIS	0.92 ±0.01	0.91 ±0.03	0.87 ±0.09	0.87 ±0.06
DERF	0.97 ±0.02	0.96 ±0.03	0.92 ±0.04	0.94 ±0.05

TABLE IV. ROOT MEAN SQUARED ERROR TRAINING ANALYSIS

Technique	20 %	40 %	60 %	80 %
LR	4.6 ±0.54	4.6 ±0.76	5.1 ±0.44	4.5 ±0.89
DT	5.5 ±0.82	4.3 ±0.41	4.5 ±0.69	4.0 ±0.79
RF	5.0 ±0.70	6.1 ±0.55	4.4 ±0.83	3.3 ±0.46
SVM	4.7 ±0.54	4.8 ±0.51	4.7 ±0.78	4.3 ±0.71
ANN	3.2 ±0.63	3.9 ±0.88	3.7 ±0.44	2.9 ±0.73
ANFIS	4.4 ±0.77	4.1 ±0.81	4.1 ±0.62	6.8 ±0.58
DERF	1.7 ±0.31	2.0 ±0.39	1.1 ±0.27	1.2 ±0.29

D. Testing Analysis

In this section, performance of the existing and the proposed machine learning methods are tested by considering the testing data. It is mainly used for evaluating the effectiveness of DERF over existing methods.

Tables V, VI, and VII depict the performance analysis of the existing and the proposed machine learning methods. It is found that DERF outperforms existing methods in terms of accuracy, coefficient of determination, and Root mean squared error, respectively.

TABLE V.   TESTING ACCURACY ANALYSIS

Dataset	20 %	40 %	60 %	80 %
LR	91.3 ±1.7	89.4 ±2.1	90.5 ±1.9	89.8 ±1.3
DT	92.6 ±1.6	90.9 ±0.8	91.9 ±1.2	90.1 ±1.5
RF	93.1 ±1.8	90.6 ±2.0	91.3 ±1.3	91.7 ±2.2
SVM	93.1 ±1.0	90.6 ±1.7	91.3 ±2.1	91.7 ±1.9
ANN	93.5 ±1.1	93.9 ±1.3	92.6 ±1.6	95.9 ±1.0
ANFIS	94.2 ±1.0	94.7 ±1.3	93.4 ±2.1	96.3 ±1.1
DERF	98.1 ±0.8	98.4 ±0.8	98.2 ±0.9	99.1 ±0.7

TABLE VI.   TESTING COEFFICIENT OF DETERMINATION ANALYSIS

Dataset	20 %	40 %	60 %	80 %
LR	0.81 ±0.08	0.84 ±0.07	0.82 ±0.07	0.88 ±0.07
DT	0.81 ±0.07	0.84 ±0.06	0.82 ±0.07	0.89 ±0.06
RF	0.82 ±0.08	0.85 ±0.07	0.83 ±0.06	0.89 ±0.06
SVM	0.83 ±0.07	0.86 ±0.08	0.84 ±0.06	0.91 ±0.05
ANN	0.84 ±0.09	0.87 ±0.04	0.85 ±0.11	0.92 ±0.06
ANFIS	0.85 ±0.10	0.88 ±0.09	0.86 ±0.11	0.93 ±0.05
DERF	0.95 ±0.04	0.96 ±0.03	0.96 ±0.02	0.98 ±0.01

TABLE VII.   ROOT MEAN SQUARED ERROR TESTING ANALYSIS

Dataset	20 %	40 %	60 %	80 %
LR	5.1 ±0.63	5.4 ±0.78	3.7 ±0.47	3.0 ±0.68
DT	3.1 ±0.49	3.7 ±0.91	4.4 ±0.96	3.8 ±1.21
RF	5.6 ±0.41	4.8 ±0.37	3.3 ±0.81	5.5 ±0.76
SVM	4.6 ±0.47	3.3 ±0.91	4.7 ±0.74	6.3 ±0.35
ANN	4.4 ±0.65	3.9 ±0.61	3.8 ±0.45	4.9 ±0.47
ANFIS	6.3 ±0.39	6.5 ±0.48	5.4 ±0.51	3.5 ±0.55
DERF	2.5 ±0.24	1.9 ±0.27	1.6 ±0.19	1.9 ±0.28

VI.   CONCLUSION

The examination of drug synergy score needs well-organized regression models to decrease the prediction errors. The main objective of this paper is to design a novel differential evolution based multinomial random forest (DERF) approach. The proposed strategy has been assessed on the data set of drug synergy and furthermore contrasted with aggressive machine learning techniques. In experimental consequences, it has been seen that mean enhancement of proposed method over competitive methods in terms of accuracy and coefficient of determination are 2.3598 % and 1.8469 %, respectively. Hence, DERF is effective for planning a estimator of a real-time drug synergy. In this work, we have not considered the use of feature selection techniques. Therefore, to improve speed and accuracy rate, we may utilize some competitive FS techniques.

REFERENCES

[1] H. Chen and J. Li, "Drugcom: Synergistic discovery of drug combinations using tensor decomposition," in *2018 IEEE International Conference on Data Mining (ICDM)*.   IEEE, 2018, pp. 899–904.

[2] H. Singh, P. S. Rana, and U. Singh, "Prediction of drug synergy in cancer using ensemble-based machine learning techniques," *Modern Physics Letters B*, vol. 32, no. 11, p. 1850132, 2018.

[3] C. Zhang and G. Yan, "Synergistic drug combinations prediction by integrating pharmacological data," *Synthetic and systems biotechnology*, vol. 4, no. 1, pp. 67–72, 2019.

[4] P. Ding, R. Yin, J. Luo, and C. K. Kwoh, "Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological and network knowledge," *IEEE journal of biomedical and health informatics*, 2018.

[5] R. Rahman and R. Pal, "A mathematical framework for analyzing drug combination toxicity for personalized medicine applications," in *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*.   IEEE, 2016, pp. 13–16.

[6] P. K. Mukherjee, S. Banerjee, and A. Kar, "Exploring synergy in ayurveda and traditional indian systems of medicine," *Synergy*, 2018.

[7] M. Shin, D. Jang, H. Nam, K. H. Lee, and D. Lee, "Predicting the absorption potential of chemical compounds through a deep learning approach," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 2, pp. 432–440, 2018.

[8] J. K. Tom, T. J. Albin, S. Manna, B. A. Moser, R. C. Steinhardt, and A. P. Esser-Kahn, "Applications of immunomodulatory immune synergies to adjuvant discovery and vaccine development," *Trends in biotechnology*, 2018.

[9] H. Li, S. Hu, N. Neamati, and Y. Guan, "Taiji: approaching experimental replicates-level accuracy for drug synergy prediction," *Bioinformatics*, vol. 10, 2018.

[10] M. Yang, M. P. Menden, P. Jaaks, J. Dry, M. Garnett, and J. Saez-Rodriguez, "Stratification and prediction of drug synergy based on target functional similarity," *bioRxiv*, p. 586123, 2019.

[11] L. Huang, D. Brunell, C. Stephan, J. Mancuso, X. Yu, B. He, T. C. Thompson, R. Zinner, J. Kim, P. Davies *et al.*, "Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction," *Bioinformatics*, 2019.

[12] C. Gilvary, J. R. Dry, and O. Elemento, "Multi-task learning predicts drug combination synergy in cells and in the clinic," *bioRxiv*, p. 576017, 2019.

[13] A. Jaiswal, B. Yadav, K. Wennerberg, and T. Aittokallio, "Integrated analysis of drug sensitivity and selectivity to predict synergistic drug combinations and target coaddictions in cancer," in *Systems Chemical Biology*.   Springer, 2019, pp. 205–217.

[14] K. Savva, M. Zachariou, A. Oulas, G. Minadakis, K. Sokratous, N. Dietis, and G. M. Spyrou, "Computational drug repurposing for neurodegenerative diseases," in *In Silico Drug Design*.   Elsevier, 2019, pp. 85–118.

[15] Z. B. Weinstein, A. Bender, and M. Cokol, "Prediction of synergistic drug combinations," *Current Opinion in Systems Biology*, vol. 4, pp. 24–28, 2017.

[16] R. Sałat and K. Sałat, "Modeling analgesic drug interactions using support vector regression: a new approach to isobolographic analysis," *Journal of pharmacological and toxicological methods*, vol. 71, pp. 95–102, 2015.

[17] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, "Predicting potential side effects of drugs by recommender methods and ensemble learning," *Neurocomputing*, vol. 173, pp. 979–987, 2016.

[18] L. Wang, Z.-H. You, X. Chen, S.-X. Xia, F. Liu, X. Yan, and Y. Zhou, "Computational methods for the prediction of drug-target interactions from drug fingerprints and protein sequences by stacked auto-encoder deep neural network," in *International Symposium on Bioinformatics Research and Applications*.   Springer, 2017, pp. 46–58.

[19] T. Turki, Z. Wei, and J. T. Wang, "Transfer learning approaches to improve drug sensitivity prediction in multiple myeloma patients," *IEEE Access*, vol. 5, pp. 7381–7393, 2017.

[20] I. F. Tsigelny, "Artificial intelligence in drug combination therapy," *Briefings in bioinformatics*, 2018.

[21] R. Hemalatha and T. Devi, "Prognosticate the drugs for multiple myeloma patients by using gene expression technique with polyclonal and monoclonal samples." *ICTACT Journal on Image & Video Processing*, vol. 8, no. 3, 2018.

[22] Y. Yuan, F. Zheng, and C.-G. Zhan, "Improved prediction of blood–brain barrier permeability through machine learning with combined use of molecular property-based descriptors and fingerprints," *The AAPS journal*, vol. 20, no. 3, p. 54, 2018.

[23] I. Olier, N. Sadawi, G. R. Bickerton, J. Vanschoren, C. Grosan, L. Soldatova, and R. D. King, "Meta-qsar: a large-scale application of meta-learning to drug design and discovery," *Machine Learning*, vol. 107, no. 1, pp. 285–311, 2018.

[24] M. E. Levine, D. J. Albers, and G. Hripcsak, "Methodological variations in lagged regression for detecting physiologic drug effects in ehr data," *Journal of biomedical informatics*, vol. 86, pp. 149–159, 2018.

[25] R. Su, X. Liu, L. Wei, and Q. Zou, "Deep-resp-forest: A deep forest model to predict anti-cancer drug response," *Methods*, 2019.

[26] W. Li, X. Yan, J. Pan, S. Liu, D. Xue, and H. Qu, "Rapid analysis of the tanreqing injection by near-infrared spectroscopy combined with least squares support vector machine and gaussian process modeling techniques," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019.

[27] R. B. Mofrad, N. S. Schoonenboom, B. M. Tijms, P. Scheltens, P. J. Visser, W. M. van der Flier, and C. E. Teunissen, "Decision tree supports the interpretation of csf biomarkers in alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 11, pp. 1–9, 2019.

[28] S. S. Bashar, M. S. Miah, A. Z. Karim, M. A. Al Mahmud, and Z. Hasan, "A machine learning approach for heart rate estimation from ppg signal using random forest regression algorithm," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2019, pp. 1–5.

[29] A. Ogunleye, Q.-G. Wang, and T. Marwala, "Integrated learning via randomized forests and localized regression with application to medical diagnosis," *IEEE Access*, vol. 7, pp. 18 727–18 733, 2019.

[30] J. Zhao, S. Gu, and A. McDermaid, "Predicting outcomes of chronic kidney disease from emr data based on random forest regression," *Mathematical biosciences*, vol. 310, pp. 24–30, 2019.

[31] Y. Feng, P. Zhang, M. Yang, Q. Li, and A. Zhang, "Short term load forecasting of offshore oil field microgrids based on da-svm," *Energy Procedia*, vol. 158, pp. 2448–2455, 2019.

[32] B. Kestenbaum, "Linear regression," in *Epidemiology and Biostatistics*. Springer, 2019, pp. 217–237.

[33] Y. Jin, H. Yu, Y. Zhang, N. Pan, and M. Guizani, "Predictive analysis in outpatients assisted by the internet of medical things," *Future Generation Computer Systems*, 2019.

[34] P. Rajalakshmy, E. Jacob, and T. J. Sharon, "Estimation of elbow joint angle from surface electromyogram signals using anfis," in *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*. Springer, 2019, pp. 247–253.

[35] L. A. Hazelden, M. J. Newman, S. Shuey, J. M. Waldfogel, and V. T. Brown, "Evaluation of the head and neck cancer patient population and the incidence of hospitalization at an academic medical center," *Journal of Oncology Pharmacy Practice*, vol. 25, no. 2, pp. 333–338, 2019.

[36] Y. Guan, "Guanlab's solution to the 2016 astrazeneca-sanger drug combination prediction dream challenge," *Synapse Repository, Synapse ID: syn5614689*, 2016.