



COVID-19 drug repurposing model based on pigeon-inspired optimizer and rough sets theory

Ibrahim Gad¹ · Mohamed Torky² · Yaseen A. M. M. Elshaier^{3,7} · Ashraf Darwish^{4,7} · Aboul Ella Hassanien^{5,6,7}

Received: 10 December 2022 / Accepted: 14 January 2024 / Published online: 22 February 2024

© The Author(s) 2024

Abstract

Discovering the most effective anti-SARS-CoV-2 drugs is the optimal solution to get back to a normal life without COVID-19. Drug repurposing, also known as drug repositioning, has become one of the most important solutions for developing new COVID-19 drugs. However, this alternative requires long-term laboratory experiments to reach the optimal drug that involves the best combination of drug features to resist the COVID-19 virus. In response to this challenge, the COVID-19 drug repurposing (C19-DR) model based on pigeon-inspired optimizer (PIO) and rough sets theory (RST) is proposed. The proposed model presents a new rough set-based feature selection technique that uses a pigeon-inspired optimizer algorithm to find and validate the optimal reduct of drug features to design an effective COVID-19 drug. Moreover, the proposed model can investigate the efficiency of multiple medications against the COVID-19 virus based on the half-maximal inhibitory concentration (IC₅₀) threshold. The effectiveness of the proposed COVID-19 drug repurposing model has been validated using a laboratory drug dataset consisting of 60 medications. The practical results show that the optimized rough set reduct of {hydrogen bonding acceptor (HBA) and number of chiral centers} is the most significant reduct that can be used to design an effective COVID-19 drug. Moreover, the proposed drug design model could verify the efficiency of a selected dataset of drug models based on evaluating the IC₅₀ metric. The verification results proved the high effectiveness of the proposed model in evaluating the predicted IC₅₀ with an accuracy of 91.4% and MSE of 0.034. These findings might be a promising solution that can assist researchers in developing and repurposing novel medications to treat COVID-19 and its new viral mutants.

Keywords COVID-19 · Drug repurposing · Feature selection · Pigeon-inspired optimizer · Rough sets theory

1 Introduction

In drug industry, drug repurposing has become a central issue for discovering new drug and vaccine models. The process of developing new drugs is time-consuming and involves a significant amount of effort to fight the viruses that are being targeted. The pharmaceutical industry has depended on drug repositioning, also known as drug repurposing, refiling, redirection, and therapeutic flipping, as a potentially effective strategy for drug discovery for decades. Medication repositioning is based on finding novel applications for current or discontinued medications with safe and effective pharmacokinetic profiles [1]. Drug repositioning entails the combination of pharmaceuticals

with varied indications, decreasing drug resistance, and the more effective treatment of severe illnesses. The effectiveness of drug repurposing has been significantly enhanced by the usage of computational methodologies such as ligand-based drug design, structure-based drug design, and more recently, the integration of artificial intelligence (AI) techniques [2–4].

The world is still suffering from the COVID-19 virus and its mutants [5]. Although all countries have applied multiple precaution procedures to block COVID-19 infection propagation and limit the number of infected people, the world is still suffering from viral mutants. One of the fastest and most effective approaches to resist the COVID-19 pandemic is drug repurposing [6]. The idea is based on the interaction between drugs and a small component of the body tissue of the receptor. To cause the biological influence, the receptor will interact with particular pharmacologic properties included in the drug. The descriptors

Ibrahim Gad and Mohamed Torky are contributed equally to this work.

Extended author information available on the last page of the article

features are components of the medicine that has important biological and chemical effects. The following routes are involved in the development of drug–receptor interactions: ion–dipole interaction, HBs, van der Waals interactions, and ionic bond formation. These are examples of reversible reactions [7]. HBs are created when electronegative atoms (Cl, F, N, S, and O) come into contact with a proton that has been bonded to one of the electronegative atoms. Several drugs have amino, hydroxyl, carbonyl, or carboxyl groups, and the receptor protein has many interchangeable hydrogen atoms that may be replaced. HBs have an important function in defining ligand (drug) activity specificity, which is essential for drug development. The drug–receptor interaction can be dissociated because HBs break down quickly.

AI is fast becoming a popular technology in medicine and pharmacy [8]. It has many medical applications, such as disease recognition and diagnosis, and end-to-end drug discovery and development. In pharmacology, AI can assist in many applications including drug discovery, dose determinations, and drug impact monitoring. Using AI-based expert systems, doctors can make the best and optimal medical diagnosis decisions. Moreover, AI aids them with drug therapy tracking, medication combinations, and the best selection of drug formularies [6].

Machine learning is an interesting approach to AI because it enables machines or software systems to make the right decisions based on previously training data. As a result, machine learning algorithms can be utilized effectively in drug repurposing and discovery processes [9, 10]. Since data gathering is usually highly dimensional and imbalanced, particularly in medical datasets, some real-world applications have been restricted in their ability to be utilized. A critical step in the development of an effective machine learning model is the feature selection methodology, which removes any unnecessary attributes that might potentially decrease the accuracy of the model. Feature selection is a technique used for picking a subset of attributes from a large set of data attributes to reduce the dimensionality of the feature space to carry out an efficient classification task [11].

On the other hand, the step of feature selection is an important aspect of the machine learning approach. The process can be performed manually by an expert or automatically by several various algorithms [12]. One of the most effective mathematical approaches that can be used for feature selection is the rough sets theory (RST). The main advantage of RST is that it preserves the qualities of the original features while deleting unnecessary information. A rough set-based feature selection is based on the concept of finding a globally minimum reduct, which is the minimal set of features from a large set of features that contain the most significant information. The current

feature selection approaches are insufficient and ineffective enough to identify a globally minimum reduct of the best minimum features. This limitation makes that RST is a good alternative for developing efficient feature selection techniques.

Although greedy strategies, often known as hill-climbing approaches, are characterized by using heuristic information based on a rough set of feature significance, they did not work well in determining the minimal reduct of features. The forward features selection method of greedy algorithms might make them effective in making decisions about candidate features based on the relevance of their characteristics. However, these algorithms did not direct the search in the optimal path to getting the minimum and optimal reduct of best features [13].

This paper introduces a significant rough set-based COVID-19 drug repurposing model. The optimized feature selection process of this model is performed using pigeon-inspired optimizer (PIO), which can determine the optimal pharmacological reducts-based features (subset) through an RST-based fitness function to design a novel COVID-19 drug model [14, 15]. The experimental results have shown that the pharmacological feature and the number of rotatable bonds were shown to be the most significant variable that accurately designs COVID-19 drugs. In addition, the optimized rough set reduct of (hydrogen bonding acceptor (HBA) and number of chiral centers) is shown to be the optimal reduct that can be selected to design a novel COVID-19 drug model. Moreover, the proposed COVID-19 drug repurposing model can be used to validate the efficiency of multiple drugs against the COVID-19 virus based on the half-maximal inhibitory concentration (IC₅₀) measure. This paper introduces a significant rough set-based COVID-19 drug repurposing model. The optimized feature selection process of this model is performed using pigeon-inspired optimizer (PIO), which can determine the optimal pharmacological reducts-based features (subset) through an RST-based fitness function to design a novel COVID-19 drug model [14, 15]. The experimental results have shown that the optimized rough set reduct of {hydrogen bonding acceptor (HBA) and the number of chiral centers} is shown to be the optimal reduct that can be selected to design a novel COVID-19 drug model. Moreover, the proposed COVID-19 drug repurposing model can be used to validate the efficiency of multiple drugs against the COVID-19 virus based on the half-maximal inhibitory concentration (IC₅₀) measure.

The main contributions of this paper are summarized as follows:

- The proposed COVID-19 drug repurposing (C19-DR) model used to discover a new medication to COVID-19 coronavirus.

- The proposed model has been designed based on pigeon-inspired optimizer and rough sets theory to design the optimal drug that involves the best reduct of pharmacological features to resist the COVID-19 coronavirus.
- The proposed model investigates the efficiency of multiple medications against the COVID-19 virus based on the half-maximal inhibitory concentration (IC50) threshold.
- The proposed model can precisely predict the usefulness of multiple medications against the COVID-19 virus based on calculating the IC50 value.
- The proposed model predicted an IC50 with an R^2 of 90%.
- The reduct of hydrogen bonding acceptor (HBA) and number of chiral centers was shown to be the optimal reduct that can be selected to design a novel COVID-19 drug.

2 Related work

Over the past decade, most research in drug repurposing has emphasized the use of classical pharmacological techniques such as molecular docking, drug signature matching, gens association, pathway mapping, and retrospective clinical analysis [16]. Only in the past 10 years have studies of drug repurposing directly addressed how artificial intelligence (AI) can be used to discover and develop various drug repurposing models [17, 18], for instance, cancer drug discovery [19], central nervous system diseases [20], and hepatocellular carcinoma [21]. There are several large cross-sectional studies, which suggest AI models for discovering COVID-19 drug designs [22].

Mohammad et al. [23] conducted a comprehensive investigation of a set of four deep learning algorithms, Hopfield network and long short-term memory (LSTM), restricted Boltzmann machine (RBM), deep Boltzmann machines (DBM), and deep belief network (DBN) for discovering COVID-19 drug repurposing model.

Delijewski et al. [24] introduced a supervised machine learning technique based on a trained dataset encoded in chemical fingerprints. The proposed model has been validated using a set of measures for evaluating the model's possibility to generalize to new chemical spaces and

contribute to designing a novel COVID-19 drug. The results clarified that the zafirlukast drug is the best repurposing model for COVID-19.

Aghdam et al. [25] proposed an informative feature selection technique based on a machine learning approach for COVID-19 drug redesign. The proposed methodology is based on identifying the major proteins relative to COVID-19 pathology as candidate drug target features through a pre-established biological network of COVID-19. The practical results produced five clusters of drugs that have the best features for treating COVID-19 coronavirus.

Behdad et al. [26] proposed a novel deep learning framework to explain how COVID-19 drug repurposing can be performed and accelerated using eight layers of an artificial intelligence model. The theoretical model can be utilized for analyzing, identifying, and predicting COVID-19 drug repurposing stages according to the author's suggestion.

Sharma et al. [27] investigated the efficiency of a proposed deep learning technique called Deep-AVPpred for discovering peptide drugs suitable for infection viral such as COVID-19. It is a novel deep learning classifier utilized for discovering antiviral peptides (AVPs), which can be used for designing antiviral drugs for use in veterinary and human medicine. The experimental results have shown that Deep-AVPpred achieved precision values of 94% and 93% on validation and test datasets, respectively. This result refers to the effectiveness of Deep-AVPpred in predicting novel AVPs for developing drug repurposing models for viral infections.

Ramadhan et al. [28] used the random forest algorithm to predict the effectiveness of eucalyptus, an herb against COVID-19 coronavirus. The experimental results clarified that some eucalyptus compounds, such as 1, 8-cineole, and alpha-terpinene, can interact with the COVID-19 protein; therefore, eucalyptus herb can be used for developing novel COVID-19 drug repurposing models.

All of the studies reviewed here support the possibility of using machine and deep learning techniques for developing COVID-19 drug repurposing models. However, most of these AI models are not new, not sufficient, and still require more improvements in their classification and regression accuracy for developing COVID-19 drug repurposing models. Therefore, proposing new approaches for COVID-19 drug repurposing is a necessary issue and still requires more research work. COVID-19 drug

repurposing based on rough set theory is a novel approach and fertile area of research [29].

The main strengths of rough sets are that they can be utilized for data classification to discover structural relationships within noisy, incomplete, and imprecise data. However, the main challenge with rough sets is rough set reduction (i.e., identifying the minimal features of datasets that can be used for classification tasks). In response to this challenge, many researchers utilized swarm intelligent algorithms for solving rough set reduction problems, for example, colony optimization [30], particle swarm optimization [31], and fish swarm algorithm [32]. Other solutions can be based on heuristic algorithms (i.e., greedy approximation) such as genetic algorithms [33]. However, these classes of algorithms may fail in detecting the minimal and optimal rough set reducts. Therefore, swarm intelligent algorithms inspired by nature are more effective in detecting rough set reducts than heuristic algorithms.

3 Theoretical foundation of mathematical models

This section provides a background about the principles of rough sets theory, and how it can be used to produce a set of reductions while doing a features selection procedure. Moreover, the principles of the methodology of pigeon-inspired optimizer are discussed to clarify how produced reduct sets of features are optimized efficiently.

3.1 Principles of rough sets theory

Rough sets theory (RST) is a mathematical discipline initially proposed by Z. Pawlak in 1982 [34]. RST can be defined as a mathematical model that specifies a special class of set theory in which the uncertainty of its elements can be controlled by two sets, named the lower and upper approximations. A detailed explanation of RST is discussed in [35].

An information system (IS) can be mathematically represented by the notation $IS = (I, A, D, \phi)$, where I represents a finite set of rows or objects in the information system. Each object corresponds to a specific entity or instance under consideration, such as patients in a medical database or customers in a sales dataset. A denotes a defined set of columns, attributes, or features in the

information system. These attributes describe different characteristics or properties of the objects in the system. Examples of attributes could include age, gender, and income.

D signifies the union of feature domains, which is represented as $D = \cup_{a \in A} D_a$. Here D_a represents the possible values or states that attribute a can have. For instance, if an attribute represents the color of an object, D_a could be the set of colors such as red, blue, and green. ϕ_a is a function that maps each object in I to a value in the corresponding feature domain D_a . In other words, $\phi_a : I \rightarrow D_a$ assigns a specific value from D_a to each object for attribute a . This function determines the specific value of attribute a for each object in the information system.

On the other hand, the decision system can be specified as $S = (I, A \cup \{d\}, D, \phi)$, where A and d represent condition attributes and target (decision) features, respectively.

Assuming that the subset $U \subseteq A$, the equivalence relation is denoted by Eq. 1.

$$R_U = \{(x_1, x_2) \in I \times I : \forall a \in U, a(x_1) = a(x_2)\} \quad (1)$$

The partition of I , created by R_U , is represented by the notation I/R_U and is denoted by Eq. 2.

$$I/R_U = \{[x] : x \in I\} \quad (2)$$

where $[x]$ is the equivalence class, i.e., $[x] = \{y \in I \mid (x, y) \in R_U\}$

The R -upper and R -lower approximations of a subset $V \subseteq I$ are defined by Eqs. 3 and 4, respectively.

$$R^+(x) = \{x \in I : [x]_R \cap V \neq \phi\} \quad (3)$$

$$R_+(x) = \{x \in I : [x]_R \subseteq V\} \quad (4)$$

A rough set of V with respect to the equivalence relation R is denoted by the order pair $(R_+(x), R^+(x))$.

The feature set A is defined as the union of two subsets: F and L ($A = F \cup L$). F represents the condition features, while L represents the decision features. The positive region includes all I rows that can be categorized into I/L classes based on the data provided in feature F . The positive region can be mathematically represented by Eq. 5.

$$POS_F(L) = \cup_{x \in I/L} R_+(x) \quad (5)$$

The degree of dependency can be used to identify rough set reducts.

$$\gamma_F(L) = \frac{\|POS_F(L)\|}{\|I\|} \tag{6}$$

where the notation $\|I\|$ represents the cardinality of set I . $\gamma_F(L)$ represents the dependency or regression quality between the set of condition features (F) and the set of decision features (L) in an information system [35].

The dependency function plays a crucial role in determining the predictive performance of a feature set and can be utilized as a feature importance metric, as shown in Eq. 7.

$$\beta_F(L) = R^2(\gamma_F(L)) \tag{7}$$

where R^2 is defined as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{8}$$

The reduction of features is crucial for improving computational efficiency, reducing noise or redundancy, and enhancing model interpretability. This reduction process involves comparing the equivalence relations of different feature sets. By comparing equivalence relations, it is possible to identify redundant or irrelevant features that do not contribute significantly to the predictive power of the decision feature. These features can then be eliminated, resulting in a reduced feature set that retains the same predictive potential. Features with high dependency values are considered more important and have a stronger influence on the performance and efficiency of predictive models.

3.2 Reducts

Reduct is a major concept of rough set theory. Reduct is defined as a minimized subset of features from the original set with keeping the accuracy of the original set. Therefore, reduct is frequently employed during the feature selection process to eliminate redundant and irrelevant features to reach the optimal decision for a specific problem.

A reduct is mathematically defined as a minimum subset M of the original feature set F such that, for a given collection of features L , $\gamma_M(L) = \gamma_F(L)$. Also in this case, the minimum subset M is defined as $\gamma_{M-\{a\}}(L) \neq \gamma_M(L) \forall a \in M$. As a result, no features in subset M may be eliminated without impacting the degree of dependency. In this case, according to this concept, the global minimum may not be the minimal subset (a reduct of smallest cardinality). There may be several reduct sets in a given dataset, and the term indicates the collection of all reducts.

$$M_{all} = \{X : X \subseteq F, \gamma_X(L) = \gamma_F(L); \gamma_{X-\{a\}}(L) \neq \gamma_X(L), \forall a \in X\} \tag{9}$$

The core is defined as the intersection of all the sets in M_{all} . The core consists of a subset of features that cannot be removed without introducing more contradictions to the representation of the dataset. A reduct with a minimum cardinality is an ideal objective for many problems, including feature selection. This process involves a careful examination and evaluation of the elements in order to identify a specific element within the reduct set denoted as $M_{min} \subseteq M_{all}$.

$$M_{min} = \{X : X \in M_{all}, \forall Y \in M_{all}, |X| \leq |Y|\} \tag{10}$$

3.3 The principle of pigeon-inspired optimizer (PIO)

The PIO is a swarm intelligence method that is influenced by nature [32]. Many researchers have used swarm methods to solve optimization challenges. Recent studies show that the PIO method can handle a large number of optimization problems, such as the planning of air robots' routes, automatic landing system, and PID design controller [33]. The main goal is to improve swarm solutions by changing the speed and location of each individual according to a mathematical equation inspired by natural swarm behavior. Algorithm 1 shows the main steps of the PIO.

Algorithm 1 Pigeon-inspired optimizer

Data: Initialize parameters (Compass Factor R , Space Dimension D , Number of Iterations NC , Np)

Result: Return the best reducts

$w_1 = 0.70, w_2 = 0.08, w_3 = 0.22$

Threshold = 0.1

Initialize X_i for each pigeon at random

Evaluate pigeons ($X_1, X_2, X_3, \dots, X_{NP}$)

$X_g = \text{Best fitness } (X_1, X_2, X_3, \dots, X_{NP})$

while $NC \geq 1$ **do**

Train and Test the model with the selected reduct

Evaluate the model using the fitness function-based rough set

$$\text{Fitness} = w_1 * \beta_C(D) + w_2 * MSE + w_3 * \frac{SF}{TF}$$

if ($X_p.\text{fitness}(\text{Threshold}) > X_g.\text{fitness}(\text{Threshold})$) **then**

$X_g = X_p$

else

$X_g = X_g$

$X_p = X_p$

Update X_p, X_g

Update velocity and path for Pigeon

$NC = NC - 1$

while $Np \geq 1$ **do**

$Np = Np/2$

Sort pigeons by the fitness function

Calculate the desired destination $X_c(t + 1)$ by Equation 14

Update Pigeon position $X_i(t + 1)$ by Equation 15

Update X_g

To demonstrate how PIO works, the pigeon i 's velocity (V_i) and location (X_i) are modified for each iteration. $V_i(t + 1)$ and $X_i(t + 1)$ are adjusted in accordance with the current iteration t^{th} , which is shown in Eqs. 11 and 12.

$$V_i(t + 1) = V_i(t) * e^{-Rt} + \text{rand}(X_g - X_i(t)) \quad (11)$$

$$X_i(t + 1) = X_i(t) + V_i(t + 1) \quad (12)$$

At each iteration t , $V_i(t)$ represents the pigeon's velocity, X_g denotes the global optimal solution, $\text{rand}()$ indicates a random number, R represents the map and compass factor, and $X_i(t)$ represents the pigeon's current position.

All pigeons are ranked according to their fitness score in the landmark operator. From Eq. 13, just 50 percent of the pigeon population (N_p) is taken into account when it comes time to determine where to place the center pigeon in each subsequent generation. However, the other pigeons change

their flight path based on the ideal position of their final destination. There are two equations involved in this process: One determines the location of a particular bird using Eq. 14, and the other updates the locations of all other birds using Eq. 15 [32].

$$N_p(t + 1) = \frac{N_p(t)}{2} \quad (13)$$

$$X_c(t + 1) = \frac{\sum X_i(t + 1) \cdot \text{Fitness}(X_i(t + 1))}{N_p \sum \text{Fitness}(X_i(t + 1))} \quad (14)$$

Where X_i denotes the current position of all pigeons, and X_c indicates the centering pigeon's location (desired destination).

$$X_i(t + 1) = X_i(t) + \text{rand}(X_c(t + 1) - X_i(t)) \quad (15)$$

4 The proposed COVID-19 drug repurposing (C19-DR) model

Based on the methodologies of the PIO algorithm and rough set theory, we explain in this section how a proposed COVID-19 drug repurposing (C19-DR) model can be designed. The core idea behind the proposed model is to use the PIO algorithm and generate a new fitness function

using rough set theory to find and validate the optimal reduct of pharmacological features to design an effective COVID-19 drug model. The integration of PIO and rough sets theory can return optimized reducts, which can then be ranked for designing a novel COVID-19 drug based on the best reduct of pharmacological features. Algorithm 2 describes the main steps of the proposed CR-19 model.

Algorithm 2 The main steps of the proposed C19-DR model

Data: Drug dataset D with p pharmacological features.
Set Threshold $\leftarrow 0.1$
Initialize score = DataFrame([Threshold, Fitness value, Xgb, R^2 , MSE, Reduct])

Step 1: Pre-processing dataset D :

- Normalize the dataset D by StandardScaler()
- Visualize the dataset D by heatmap
- Split D to train D_{train} and test D_{test}

Step 2: Initialize PIO parameters: (Compass Factor R , Space Dimension d , Number of Iterations NC , population size Np)

Step 3: Initialize population $P = (X_1, X_2, X_3, \dots, X_{NP})$

Step 4: Pigeon Inspired Optimizer

while $NC \geq 1$ **do**

```

    Xg, Reduct = Apply PIO (Algorithm 1)
    if Xg.fitness(Threshold) > Xp.fitness(Threshold) then
        score.append(Threshold, Fitness value, Xg, R2, MSE, Reduct)
        /* finds a rough set reduct that maximizes the
        fitness function in the current iteration. */
    else
        ⊥ Continue
    plot R2(Actual, Predicted)
    plot relative importance(Reduct)
    NC = NC - 1

```

Step 5: The best rough set reduct

- optimal reduct $R^{opt} = \text{score.sort_values}([\text{Fitness value}, R^2])$.

Step 6: Validation new drugs

- $X_{val} = D_{val}[D.columns[R^{opt}]]$
- model = GradientBoostingRegressor()
- model = model.fit()
- $D_{val}["Pred_IC50"] = \text{model.predict}(X_{val})$
- **if** $D_{val}["Pred_IC50"] \leq 5$ **then**
 - ⊥ $D_{val}["Decision"] = \text{"More effective"}$
 - else**
 - ⊥ $D_{val}["Decision"] = \text{"Less effective"}$

Output: The best reduct R^{opt}

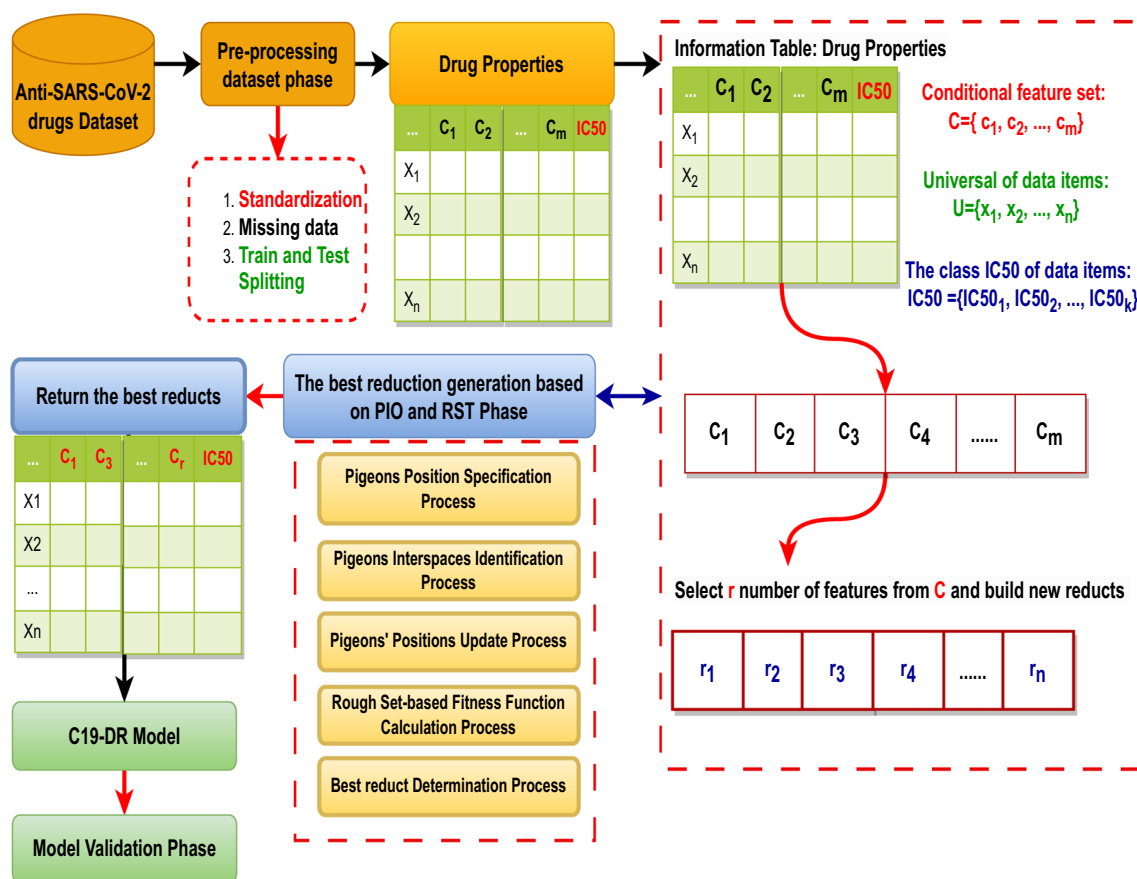


Fig. 1 The general architecture of the proposed COVID-19 drug repurposing model

Figure 1 demonstrates the general architecture that describes the main phases of the proposed model. The main phases of the proposed CR-19 model are described in detail as follows.

4.1 Pre-processing phase

The main objective of the data preprocessing phase is to prepare the drug dataset, which involves pharmacological features. Therefore, the proposed model can be efficiently trained and evaluated. This phase includes three main steps, namely: (a) Standardization: Dataset standardization involves rescaling all features to a common scale and distribution while preserving the original variables' ranges and variances. (b) Handling missing data: To train models on a consistent entire feature space, it is essential to handle missing data properly. To solve the issue of missing data, the mean method is used to fill in any dataset attributes that have missing values. (c) The training and testing splitting: The preprocessed data are randomly split into two distinct training and test subsets using common rules, such as 70%/30% or 80%/20%.

4.2 Optimal reduct generation phase

The primary objective of this study was to identify the optimal drug that involves the best combination of pharmacological features to effectively resist the COVID-19 virus. In this phase, the generation of optimal reducts is done through the following processes.

4.2.1 Pigeons position specification process

Optimizing feature selection process using PIO requires firstly to specifies pigeons positions before executing PIO method. In this process, the proposed model specifies the pigeon (solution) using a vector with a certain length N (N represents the total number of attributes). Position and velocity vectors were initialized randomly from the range $[0, 1]$. The position of the pigeon is indicated by strings of N binary bits. In this case, each bit stands for a feature, and the value of "1" means that the corresponding attribute is picked, whereas a value of "0" implies that it is not. Each position corresponds to a subset of features.

A sigmoidal function is used in Eq. 16 to transform the velocity of each pigeon $V_i(t)$ into binary form during

iteration t . Equation 17 modifies the location of each pigeon based on the sigmoid function value and the uniform random number r , which ranges from 0 to 1. After the velocities are transferred using the sigmoid function, the locations are adjusted to reflect the new values.

$$s(v_i) = \frac{1}{1 + \exp\left(\frac{v_i}{2}\right)} \quad (16)$$

$$X_{(i,p)}(t) = \begin{cases} 1, & s(v_i) > r \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

4.2.2 Pigeons interspaces identification process

The second phase in optimizing pharmacological feature selection using PIO is pigeons interspace identification. It is possible to compute the distances between two pigeons using what is called the Hamming distance. It is a mathematical metric for comparing two binary data strings. This metric can be imported here to identify the distances between pigeons in PIO algorithm. For example, the Hamming distance between X and Y , which represents the positions of two different pigeons, is the number of “1’s” in the binary bit string $X \oplus Y$.

4.2.3 Pigeons’ positions update process

This phase represents a real step toward the optimal solution of PIO algorithm in a given features selection problem. Every iteration of PIO algorithm begins with a random starting location for each pigeon in the swarm. A pigeon aims to move one step forward in the problem space based on the behaviors of searching, swarming, and continuing to follow the optimal solution. Thus, to assess these three behaviors, a fitness function is utilized. The pigeon with the best fitness value is selected to change the other pigeons’ positions.

4.2.4 Rough set-based fitness function calculation process

The fitness function is the most important technique used to get the optimal solution in a given optimization problem. In this phase, a new fitness function is proposed based on rough set theory to optimize the produced reducts of selected features. Fitness function is applied to each reduct to compute its fitness value. In particular, three components determine this function: the number of features in the reduct, the coefficient of determination (R^2) or performance of the regression model ($\beta_C(D)$), and the mean square error (MSE). The mathematical definition of the fitness function is stated in the following Eq. 18:

$$\text{Fitness} = w_1 * \beta_C(D) + w_2 * \text{MSE} + w_3 * \frac{SF}{TF} \quad (18)$$

where $\beta_C(D)$ is the performance of the regression model (also called dependency). SF represents the number of “1’s” in a binary bit position that denotes the number of selected features. The term TF refers to the total number of features. w_1, w_2 , and w_3 are three weights that correspond to the significance of regression performance, MSE, and reduct length, $w_1 \in [0, 1]$, and satisfy the following condition: $w_1 + w_2 + w_3 = 1$.

The importance of a feature in the gradient boosting model is equal to the sum of the importance’s of the corresponding nodes. It would thus be completely consistent to describe the significance of a set of features as the sum of the importance of all the related nodes if there is such a set. And the latter is precisely equal to the total of the importance of the various features. The weight function for each reduct is defined as follows:

$$\text{Reduct_weight} = \sum_{fi \in TF} \text{importance}(fi) \cdot \frac{TF - SF}{TF} \quad (19)$$

Where TF denotes the total number of features, and SF represents the number of selected features, and Reduct_weight represents the calculated weight assigned to a specific reduct. The function $\text{importance}(fi)$ corresponds to the measure of significance or relevance attributed to each fi feature. Moreover, the condition $\text{importance}(fi) > \text{threshold}$ implies that only features with an importance value exceeding a predefined threshold are considered.

4.2.5 Best reduct determination process

This is the last phase of the proposed pharmacological features selection process. In this phase, a PIO finds a rough set reduct that maximizes the fitness function in the current iteration. This indicates that in this iteration, the PIO finds its local best solution to the problem. After all pigeons have been finished, a new iteration will begin by the same way. The stop condition of PIO algorithm is either reaching the maximum number of iterations or finding the minimal reduct that consists of the best selected pharmacological features to design a novel COVID-19 drug model.

4.3 Validation of drugs phase

This phase aims to evaluate the effectiveness of the C19-DR model to determine its prediction ability on untested drug data. The C19-DR model’s performance is evaluated using a number of reducts that have different numbers of pharmacological features. To further assess its

generalizability, the C19-DR model's ability to predict the IC50 value of a newly tested dataset is also evaluated. The performance of the C19-DR model is measured using standard metrics, including the coefficient of determination (R^2) and mean squared error (MSE). Reducts generated through the C19-DR framework are evaluated and compared based on these performance metrics. Finally, after extensive validation, the proposed model can be utilized as a recommendation system to accurately predict the drug for COVID-19.

5 Experimental results and evaluations

This section illustrates the performance and validation results of the proposed C19-DR model in discovering a novel COVID-19 drug based on PIO algorithm and rough set theory. The model was implemented using Python programming language and the scikit-learn libraries. Moreover, the experimentation was conducted on a Google Colab cloud platform equipped with a 2.6-GHz processor and 32 GB of memory. To provide a comprehensive overview, this section is divided into three subsections. Firstly, Subsect. 5.1 focuses on the analysis and visualization of the dataset used in the study. Secondly, Subsect. 5.2 presents the evaluation metrics employed to assess the performance of the proposed model. Lastly, Subsect. 5.3 presents the analysis of the experimental results obtained from the model.

5.1 Drug dataset description

This section presents how the dataset is constructed from publicly available anti-SARS-CoV-2 drugs. A data collection was created for several COVID-19 medications. Molecular properties and Lipinski's rule (Rule of 5) were used to calculate the characteristics of these medications. They all have distinct descriptions on which the testing trials were performed. In addition, the data describe the medications' original pharmacological functions. Table 1 displays a few samples from the used drug dataset which consists of nine pharmacological features [14]. The

attributes of the drug dataset are: hydrogen bonding donor (HBD), hydrogen bonding acceptor (HBA), calculated logP (CLog P) as a lipophilicity metric, number of rings, molecular weight (Mwt), number of non-hydrogen atom (NHA) also known as number of heavy atoms, topological surface area (tPSA), number of chiral center, number of rotatable bonds, and IC50%.

There are a variety of statistical methods available for determining a correlation value between dataset features. The most widely used is Pearson correlation coefficient (PCC). The PCC is used to evaluate the correlation between two numerical variables in order to determine the relationship that exists between those variables. The PCC value ranges from -1 to +1, where the value of 0 means that there is no correlation and a value of ± 1 means that there is a perfect positive/negative correlation. On a dataset with many attributes, the correlation values between attributes (i.e., features) can be visualized in a correlation matrix. The correlation matrix is a symmetric matrix with all diagonal elements equal to +1. For instance, The created correlation matrix in Fig. 2 shows that there is a strong positive correlation between Mwt-feature and no. of NHA-feature with correlation value 0.94. On the other hand, there is a high negative correlation between tPSA-feature and clog P-feature with a correlation value -0.72 as shown in Fig. 2.

5.2 Evaluation metrics

Optimization models, such as the C19-DR model, are typically assessed using three widely used metrics: mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R^2). These metrics play a crucial role in evaluating the performance and convergence of optimization algorithms, enabling the determination of the best results. The utilization of MAE and MSE allows for a comprehensive analysis of the optimization outcomes, providing insights into the accuracy and precision of the model's predictions. Equations 20–22 describe how MAE, MSE, and R^2 can be calculated, respectively.

Table 1 A sample of chemical descriptors for anti-SARS-CoV-2 drugs approved by the FDA

Name	Class	Mwt	HBD	HBA	CLog P	No. of rings	No. of NHA	tPSA	No. of chiral center	No. of rotatable bonds	IC50(μ M)
Remdesivir	Antiviral	602	4	13	1.24929	4	33	201.32	1	13	1.3
Lopinavir	Antiviral	628	4	5	6.0946	4	57	120	4	15	9.1
Terconazole	Antifungal	532	0	8	4.8136	5	27	62.13	1	8	11.92
Amoxicillin	Antibiotic	365	4	6	-1.8715	3	25	132.96	4	4	16.12

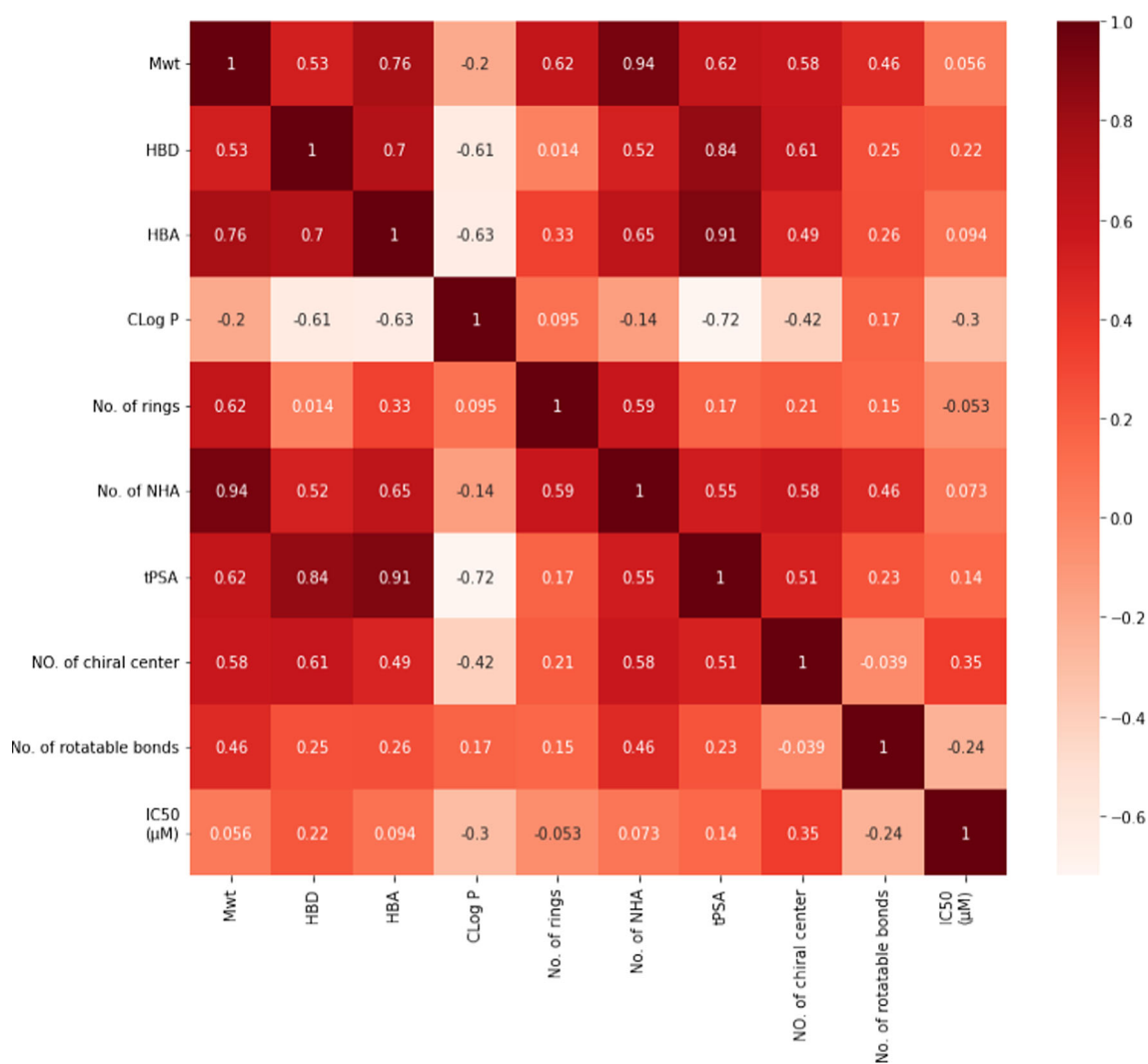


Fig. 2 The correlation matrix of the dataset

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\|. \quad (20)$$

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

where y_i is the i^{th} actual value, \hat{y}_i is the i^{th} predicted value, \bar{y} is the mean of observed values, and n is the number of observed values.

5.3 Results analysis

The primary objective of this study was to identify the optimal drug that involves the best combination of pharmacological features to effectively resist the COVID-19

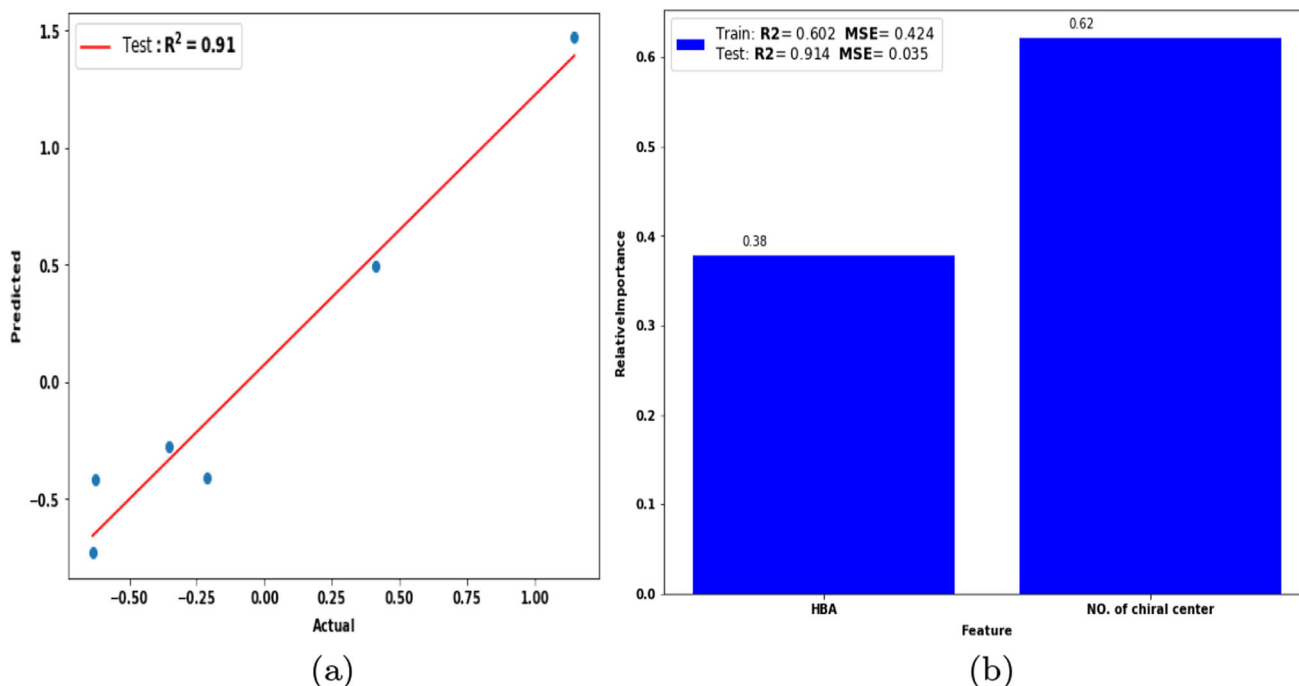
virus. In response to this challenge, the COVID-19 drug repurposing (C19-DR) model based on PIO and RST is proposed. The optimization function of the proposed C19-DR has been validated using the anti-SARS-CoV-2 drugs dataset developed in the laboratory. Moreover, the gradient boosting algorithm has been utilized for selecting the most important features in the generated reducts. To assess the optimization function of C19-DR, the measures R^2 and MSE with 10-fold cross-validation were evaluated.

The pharmacological features, which can be optimized, are hydrogen bonding acceptor (HBA), hydrogen bonding donor (HBD), calculated logP (CLog P) as a lipophilicity metric, number of rings, molecular weight (Mwt), number of non-hydrogen atom (NHA) also known as several heavy atoms, topological surface area (tPSA), number of rotatable bonds, number of chiral centers, and IC50.

Although drug dataset may contain many features that have high impacts while designing medications, few

Table 2 The minimal reducts generated by the C19-DR model

Name	R^2	Reduct weight	Reduct
Reduct1	1	0.778	[No. of chiral center, No. of rotatable bonds]
Reduct2	1	0.778	[HBA, No. of chiral center]
Reduct3	1	0.889	[No. of chiral center]
Reduct4	1	0.889	[No. of NHA]

**Fig. 3** The best dual reduct of pharmacological features used to design a novel COVID-19 drug. (a) R^2 results and (b) relative importance of the best two features

features are required for developing an effective and optimal regression model for generating drug reducts. A subset of such features is selected to increase the value of R^2 while simultaneously decreasing the value of MSE . After selecting the optimal reduct using the proposed feature selection approach, the modified dataset is used to train the selected regression model.

Table 2 shows the minimal reducts generated by the C19-DR methodology and the corresponding reduct weight and R^2 values. As a result of the dimension reduction, the generated reducts demonstrate the most important pharmacological features that can be used in designing a novel COVID-19 drug.

In addition, the results of generating different reduct sizes of the optimal features that can be used for repurposing COVID-19 drug are shown in Figs. 3, 4, 5, and 6.

Figure 3 depicts that the best dual reduct consists of two pharmacological features, number of HBA, and number of chiral center. Moreover, this combination of the two pharmacological features showed the highest R^2 value, 0.91, the highest stability value of 0.971, and the least MSE

of 0.03. This means that the two selected features hold substantial significance in the development of a novel COVID-19 medication.

Figure 4 depicts that the best triple reduct which consists of the best three pharmacological features can be combined and used for designing a novel COVID-19 drug model. This combination showed testing results, R^2 of 0.40, stability of 0.74, and MSE of 0.24. Although, this reduct clarified good training results, the testing results show that it is not recommended to be used for designing a novel COVID-19 medication.

Figure 5 shows that the best quadruple reduct which consists of the best four pharmacological features can be used for developing a novel COVID-19 drug model. This combination depicted testing results R^2 of 0.79, stability of 0.79, and MSE of 0.08. Although, this reduct clarified good training results and somewhat acceptable testing results, it is not the optimal solution for designing a novel COVID-19 medication model; hence, this combination of pharmacological features may not be recommended to use.

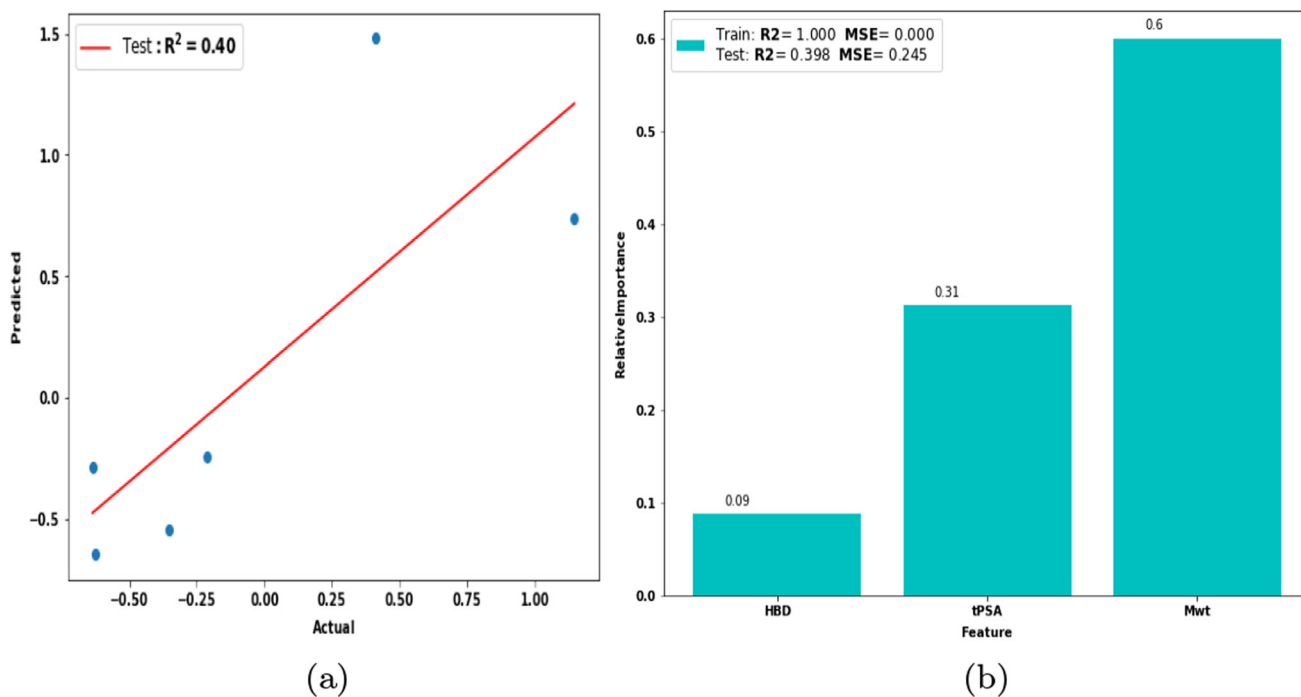


Fig. 4 The best triple reduct of pharmacological features used to design a novel COVID-19 drug. (a) R^2 results and (b) relative importance of the best three features

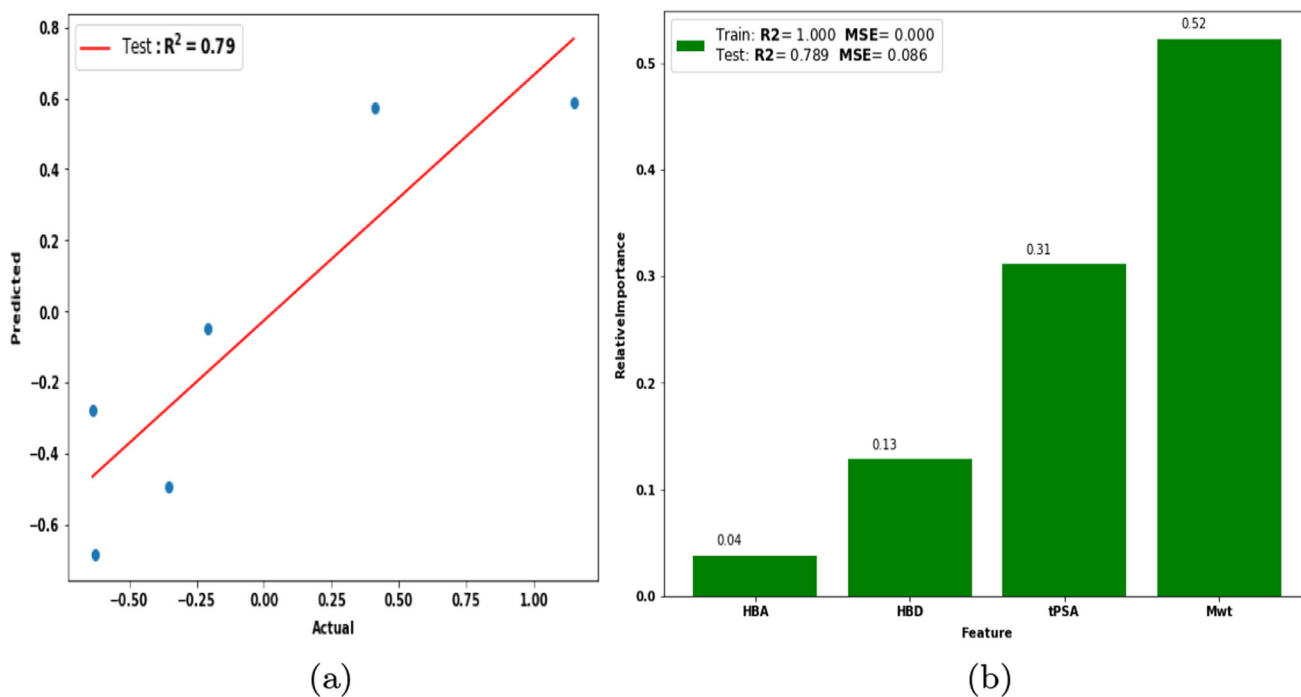


Fig. 5 The best quadruple reduct of pharmacological features used to design a novel COVID-19 drug. (a) R^2 results and (b) relative importance of the best four features

Increasing reduct size with more than four features produced bad testing results. For instance, Fig. 6 depicts that the best reduct which consists of six pharmacological features can be used to design a COVID-19 drug model.

The testing results showed that R^2 is -0.54 , stability of 0.82 , and MSE of 0.62 . Although, this reduct clarified good training results, it is failed in testing results; hence, this combination of pharmacological features and all reducts

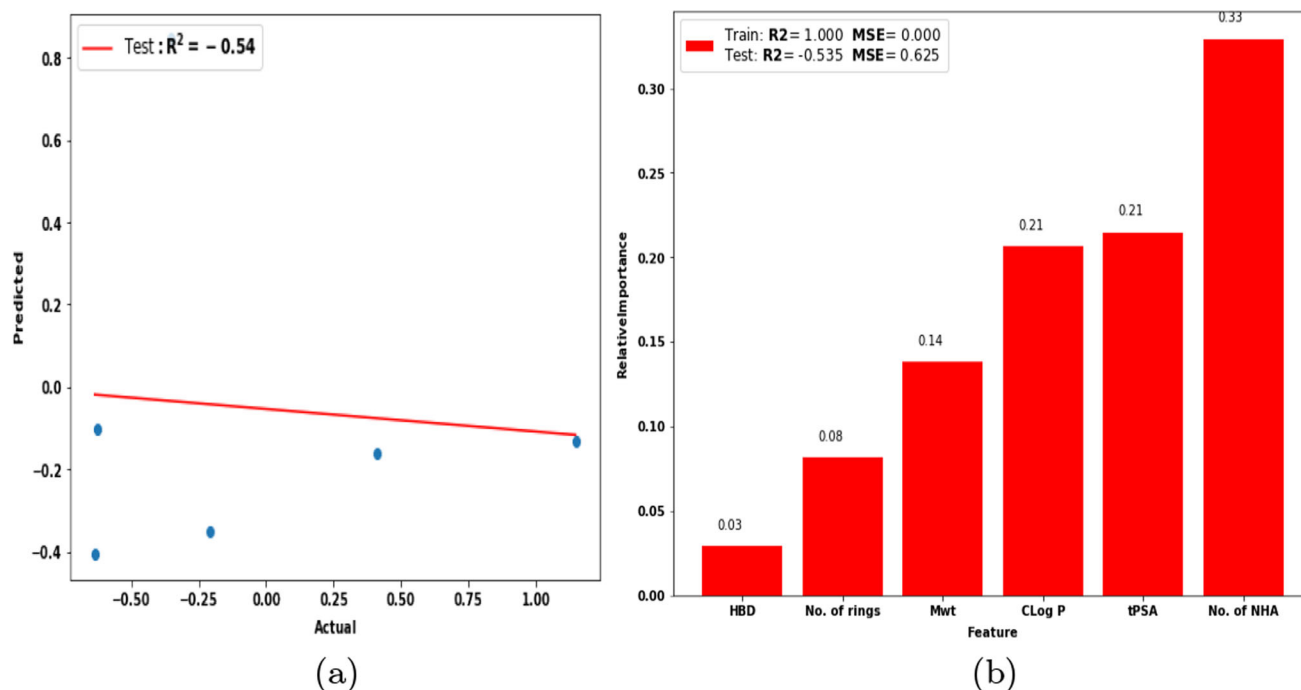


Fig. 6 The best hexagonal reduct of pharmacological features used to design a novel COVID-19 drug. (a) R^2 results and (b) relative importance of the best six features

Table 3 Training and testing results of the proposed C19-DR model in detecting all reducts can be used to design a novel COVID-19 drug model

Iteration	Train			Test			Reduct	Weight	Cardinality
	R^2	MSE	MAE	R^2	MSE	MAE			
9	1.0	0.0	0.0002	0.789	0.086	0.24	[Mwt, HBD, HBA, tPSA]	0.535	4
3	1.0	0.0	0.0004	0.398	0.245	0.346	[Mwt, HBD, tPSA]	0.608	3
7	1.0	0.0	0.0004	0.311	0.281	0.360	[HBA, No. of NHA, tPSA, NO. of chiral center]	0.514	4
2	1.0	0.0	0.0001	0.285	0.291	0.449	[Mwt, HBD, HBA, CLog P, No. of NHA, tPSA]	0.320	6
4	1.0	0.0	0.0001	0.263	0.299	0.432	[Mwt, HBA, No. of NHA, tPSA]	0.539	4
..
7	0.974	0.028	0.082	-3.246	1.728	0.891	[HBD, No. of rings, No. of chiral center, No...	0.556	4
7	0.903	0.104	0.089	-3.428	1.802	0.815	[Mwt]	0.889	1
6	0.602	0.424	0.323	0.914	0.035	0.164	[HBA, No. of chiral center]	0.885	2
2	0.538	0.492	0.391	-0.545	0.629	0.502	[No. of NHA]	0.889	1
4	0.537	0.494	0.359	-0.167	0.475	0.371	[HBA, No. of rotatable bonds]	0.778	2

that have cardinality more than four features are not recommended to use in developing a novel COVID-19 drug model.

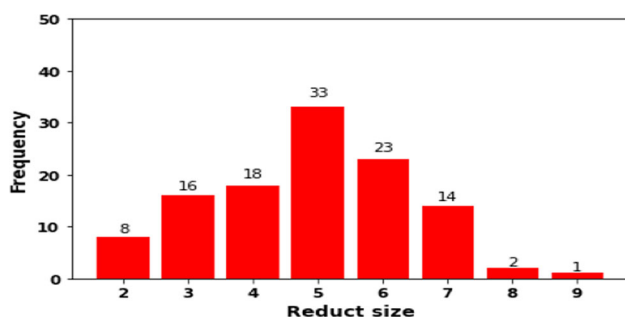
Table 3 summarizes the obtained training and testing results while simulating the methodology of the proposed C19-DR model for producing all reduct sizes can be used to design a novel COVID-19 drug model. It is clear that the dual reduct [HBA, no. of chiral center] is the most significant reduct that involves the most important

pharmacological features in designing a novel COVID-19 drug model.

Table 4 summarizes the results of the best obtained reducts (with cardinality less than or equal 4 features) that can be used effectively in designing a novel COVID-19 drug model while applying the proposed model C19-DR across 10 iterations to obtain local and global solutions by the proposed optimization methodology.

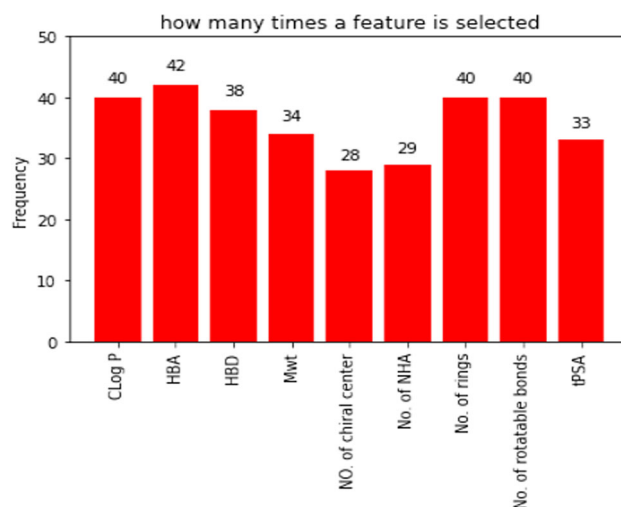
Table 4 The minimal reduct cardinalities produced by 10 iterations of applying the optimization method of the C10-DR model

Iteration	Stability value	Xgb			Xpg		
		R^2	MSE	Reduct	R^2	MSE	Reduct
0	0.847	1.000	0.000	[0, 2, 3]	1.000	0.000	[3, 4, 5]
1	0.847	1.000	0.000	[0, 2, 3]	1.000	0.000	[0, 4, 5]
2	0.871	1.000	0.000	[3, 8]	1.000	0.000	[3, 8]
3	0.871	1.000	0.000	[3, 8]	1.000	0.000	[3, 7, 8]
4	0.871	1.000	0.000	[3, 8]	0.989	0.001	[1, 2, 6]
5	0.871	1.000	0.000	[3, 6]	1.000	0.000	[3, 6]
6	0.871	1.000	0.000	[0, 3]	1.000	0.000	[0, 3]
7	0.871	1.000	0.000	[0, 3]	0.999	0.000	[0, 8]
8	0.871	1.000	0.000	[0, 3]	1.000	0.000	[2, 3, 5, 7]
9	0.871	1.000	0.000	[0, 3]	1.000	0.000	[2, 5, 6]

**Fig. 7** The frequencies for each reduct size

The frequencies for each reduct obtained by applying the proposed C19-DR models are shown in Fig. 7. Moreover, the frequencies of appearing each pharmacological feature in different reducts are depicted in Fig. 8. These results prove that the high-frequency value of a reduct or a pharmacological feature does not represent an indicator to the relative importance of that reduct or feature in developing a novel COVID-19 drug model. Whereas the optimal combination of features is based on testing results of evaluating R^2 , stability, and mean square error (MSE).

Finally, the proposed C19-DR model was utilized to investigate the IC₅₀ of the test dataset to verify the efficiency of some tested drugs while simulating the proposed methodology of C19-DR model. Each produced reduct of the drug's features has been evaluated at each trial in order to examine its influence on the expected IC₅₀ value. Table 5 shows the verification results of the tested dataset that consists of 10 drug models. The predicted IC₅₀ values are compared with the actual observed values to evaluate the performance and suitability of a tested drug model for COVID-19. It is known that the more effective drugs must have IC₅₀ value less than 5. The obtained results in Table 5 show very close predicted IC₅₀ values with the real ones while applying the proposed C19-DR on all tested drug models except with the "Remdesivir." This is the only case

**Fig. 8** How many times a feature is selected

with which the proposed model failed to correctly predict its IC₅₀ value.

6 Discussion

The existing literature on the development of optimal COVID-19 drug models using AI techniques is very little [24, 25, 27]. The current study was designed to determine the efficiency of integrating pigeon-inspired optimizer and rough sets theory to design a novel COVID-19 drug repurposing (C19-DR) model. On the question of getting the optimal minimal reduct of pharmacological features that can be used to design an effective COVID-19 drug model, the proposed C19-DR has been applied and validated on a laboratory dataset consisting of 60 drugs and nine pharmacological features. The most obvious finding to emerge from this investigation is that the dual reduct consisting of [hydrogen bonding acceptor (HBA) and number of chiral centers] showed the highest R^2 value,

Table 5 The IC50% verification results using the proposed C19-DR model

	Name	Real IC50	Predicted IC50	Decision
0	Remdesivir	1.30	12.097	Less effective
1	Lopinavir	9.10	9.099	Less effective
2	Ivermectin	2.00	2.000	More effective
3	Amodiaquine dihydrochloride	2.59	2.589	More effective
4	Amodiaquine hydrochloride	2.36	2.360	More effective
5	Benzotropine mesylate	13.80	13.799	Less effective
6	Chloroquine phosphate	42.03	42.029	Less effective
7	Chlorpromazine hydrochloride	3.14	3.139	More effective
8	Clomipramine hydrochloride	5.63	5.630	Less effective
9	Fluphenazine dihydrochloride	6.36	6.360	Less effective
10	Fluspirilene	3.16	3.160	More effective

0.91, the highest reduct's stability value of 0.971, and the least mean square error of 0.03 as shown in Fig. 3. This finding means that the two selected features are the most significant features that can be used to develop an effective COVID-19 medication model.

The analysis of the performance of C19-DR models clarified multiple reducts with different cardinalities of pharmacological features, which are suggested by the C19-DR to use in developing the COVID-19 medication model. However, the obtained results clarified that the higher number of features of a reduct, the lower the results of testing accuracy of R^2 and MSE.

Surprisingly, the results clarified that there is no impact on the frequency of a feature/reduct on the effectiveness of designing a novel COVID-19 drug model. Whereas, the trade-off among R^2 , reduct's stability, and MSE results are the most important factors. Therefore, Figs. 4, 5, and 6 did not present the optimal reducts in developing an effective COVID-19 drug model.

Another important finding was that investigating the IC50 metric of a tested dataset consists of 10 drug models. The proposed C19-DR model provided excellent training and testing results in evaluating the predicted IC50 values. As depicted previously in Table 5, the performance of the C19-DR model shows very close results in evaluating the predicted IC50 values of various 10 drug models in the tested dataset. The proposed model failed only in evaluating the predicted IC50 value of "Remdesivir." These results interpret the efficiency of the C19-DR model to be used as a recommendation system that can precisely predict

the efficiency of a given drug to COVID-19 coronavirus with an accuracy of 90%.

The proposed C19-DR model has been compared with the support vector regression (SVR) model, using the 10-fold cross-validation (CV) approach to predict the IC50 value as summarized in Table 6. The performance evaluation of the models in this study is conducted using two commonly employed metrics, namely, R^2 and MSE. The comparison results indicate that the proposed C19-DR model outperforms the other models, exhibiting the highest level of performance in accurately predicting and verifying the IC50 values.

Overall, these results must be interpreted with caution since the pigeon optimizer may constitute different reducts of features with different simulation experiments according to the configuration of the search space of the PIO technique. This indeed may impact the accuracy of obtained results in each simulation round. However, this problem does not call into question the integrity of the system design and implementation. According to these findings, we can infer that the efficiency and stability of the used optimizer and the size of the dataset are major factors on which the accuracy of getting the optimal solution results depends. This finding, while preliminary, suggests that integrating swarm intelligence optimization technique and using rough set theory to formulate novel fitness functions used by optimizers can produce an effective drug repurposing model. Further work is required to establish the viability of using other optimization techniques and their applications on multiple drugs detest discovering new medications for many chronic diseases, such as asthma, bowel syndrome, cystic fibrosis, cancer, heart disease, and lastly, monkeypox.

Table 6 Comparison results with the other models

Model	R^2	MSE
Support vector regression (SVR) model [14]	0.58	18.963
COVID-19 drug repurposing (C19-DR) model	0.914	0.034

7 Conclusion and future work

The present study was designed to examine the effectiveness of using a proposed COVID-19 drug repurposing (C19-DR) model to discover a new medication for the COVID-19 coronavirus. The proposed model has been designed based on pigeon-inspired optimizer and rough set theory to design the optimal drug that involves the best combination (i.e., reduct set) of pharmacological features to resist COVID-19 coronavirus through a novel feature selection methodology. The effectiveness of the proposed COVID-19 drug repurposing model has been implemented and validated using a laboratory dataset consisting of 60 medications. The investigation of the proposed model has shown interesting findings. One of the notable outcomes derived from this study is that the reduct of {hydrogen bonding acceptor (HBA) and number of chiral centers} was shown to be the optimal reduct that can be selected to design a novel COVID-19 drug. The second major finding was that the proposed model can precisely predict the usefulness of multiple medications against the COVID-19 virus based on calculating the IC50 value. Overall, this study strengthens the idea of depending on the rough set theory in producing effective fitness functions that can be used with various optimizers to design invented drug repurposing models. Before this study, researchers were not investigating the evidence of using the rough set for optimizing drug repurposing models.

In a future study, the fitness function and the weight function will also be important variables in the C19-DR model for feature reduction, and they will need to be developed even more to be effective. Parallel algorithms can be used to accelerate the computation of the reducts for huge datasets, which can help to reduce overall processing time. The performance of the model may be further enhanced by increasing the amount of data available or by including hyperparameter methods in the model's construction. These findings might be a promising solution that can assist researchers in developing and repurposing novel medications to treat COVID-19 and its new viral mutants. We believe that this model will be helpful and valuable for any new infectious disease.

Author contribution YAM helped in data curation and formal analysis. IG and MT helped in formal analysis, investigation, methodology, and writing original draft. IG worked in software. AD contributed to writing, review, and editing and discussion. AEH helped in conceptualization, idea, and supervision. All authors are reviewed the final version.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Su M, Hu R, Tang T, Tang W, Huang C (2023) Review of the correlation between Chinese medicine and intestinal microbiota on the efficacy of diabetes mellitus. *Front Endocrinol* 13:1085092. <https://doi.org/10.3389/fendo.2022.1085092>
- Prasad K, Kumar V (2021) Artificial intelligence-driven drug repurposing and structural biology for SARS-CoV-2. *Current Res Pharmacol Drug Discov.* 2:100042. <https://doi.org/10.1016/j.crphar.2021.100042>
- Prakash O, Khan F (2023) Drug repurposing for COVID-19 using computational methods. Bentham Science Publishers, Sharjah, pp 161–185
- Maji S, Badavath VN, Ganguly S (2023) Drug repurposing and computational drug discovery for viral infections and coronavirus disease-2019 (COVID-19). Apple Academic Press, Cambridge, pp 59–76
- Zhao Y, Hu M, Jin Y, Chen F, Wang X, Wang B, Yue J, Ren H (2023) Predicting the transmission trend of respiratory viruses in new regions via geospatial similarity learning. *Int J Appl Earth Obs Geoinf* 125:103559. <https://doi.org/10.1016/j.jag.2023.103559>
- Rodrigues L, Bento Cunha R, Vassilevskaia T, Viveiros M, Cunha C (2022) Drug repurposing for COVID-19: a review and a novel strategy to identify new targets and potential drug candidates. *Molecules* 27(9):2723
- Suvarna B (2011) Drug-receptor interactions. *Kathmandu Univ Med J* 9(3):203–207
- Krishnamurthy A, Goel P (2022) Artificial intelligence-based drug screening and drug repositioning tools and their application in the present scenario. In: Parihar A, Khan R, Kumar A, Kaushik AK, Gohel H (eds) Computational approaches for novel therapeutic and diagnostic designing to mitigate SARS-CoV-2 infection. Academic Press, Cambridge, pp 379–398
- Ahmed F, Lee JW, Samantasinghar A, Kim YS, Kim KH, Kang IS, Memon FH, Lim JH, Choi KH (2022) Speropredictor: an integrated machine learning and molecular docking-based drug repurposing framework with use case of COVID-19. *Front Public Health* 10:902123. <https://doi.org/10.3389/fpubh.2022.902123>

10. Tripathi L, Kumar P, Swain K, Pattnaik S (2022) Drug repurposing based on machine learning. Wiley, New York
11. Watt J, Borhani R, Katsaggelos AK (2020) Feature engineering and selection. Machine learning refined: foundations, algorithms, and applications. Cambridge University Press, Cambridge, pp 237–272
12. Tang X, Dai Y, Xiang Y (2019) Feature selection based on feature interactions with application to text categorization. *Expert Syst Appl* 120:207–216. <https://doi.org/10.1016/j.eswa.2018.11.018>
13. Adamczyk M (2014) Parallel feature selection algorithm based on rough sets and particle swarm optimization. In: Proceedings of the 2014 federated conference on computer science and information systems, IEEE. <https://doi.org/10.15439/2014f389>
14. Abdelhady AS, ElShaier YAMM, Refaey MS, Elmasry AE, Hassanien AE (2021) Intelligent drug descriptors analysis: toward COVID-19 drug repurposing. *Medical informatics and bioimaging using artificial intelligence: studies in computational intelligence*, vol 1005. Springer, Cham, pp 173–191
15. Hassanien AE, Bhatnagar R, Snášel V, Shams MY (2022) Medical informatics and bioimaging using artificial intelligence: challenges, issues, innovations and recent developments. Springer, Cham
16. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guillems T, Latimer J, McNamee C, Norris A, Sanssou P, Cavalla D, Pirmohamed M (2018) Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 18(1):41–58. <https://doi.org/10.1038/nrd.2018.168>
17. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK (2021) Artificial intelligence in drug discovery and development. *Drug Discov Today* 26(1):80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>
18. Xiao L, Zhang Y (2023) Ai-driven smart pharmacology. *Intell Pharm* 1(4):179–182. <https://doi.org/10.1016/j.ipha.2023.08.008>
19. Workman P, Antolin AA, Al-Lazikani B (2019) Transforming cancer drug discovery with big data and AI. *Expert Opin Drug Discov* 14(11):1089–1095. <https://doi.org/10.1080/17460441.2019.1637414>
20. Vatansever S, Schlessinger A, Wacker D, Ümit Kaniskan H, Jin J, Zhou M-M, Zhang B (2020) Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: state-of-the-arts and future directions. *Med Res Rev* 41(3):1427–1473. <https://doi.org/10.1002/med.21764>
21. Chen B, Garmire L, Calvisi DF, Chua M-S, Kelley RK, Chen X (2020) Publisher correction: harnessing big ‘omics’ data and AI for drug discovery in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol*. <https://doi.org/10.1038/s41575-020-0288-6>
22. Elkashlan M, Ahmad RM, Hajar M, Al Jasmi F, Corchado JM, Nasarudin NA, Mohamad MS (2023) A review of SARS-COV-2 drug repurposing: databases and machine learning models. *Front Pharmacol* 14:1182465. <https://doi.org/10.3389/fphar.2023.1182465>
23. Jamshidi MB, Lalbakhsh A, Talla J, Peroutka Z, Roshani S, Matousek V, Roshani S, Mirmozafari M, Malek Z, Spada LL, Sabet A, Dehghani M, Jamshidi M, Honari MM, Hadjilooei F, Jamshidi A, Lalbakhsh P, Hashemi-Dezaki H, Ahmadi S, Lotfi S (2021) Deep learning techniques and COVID-19 drug discovery: fundamentals, state-of-the-art and future directions. *Studies in systems decision and control*. Springer, Cham, pp 9–31
24. Delijewski M, Haneczok J (2021) AI drug discovery screening for COVID-19 reveals Zafirlukast as a repurposing candidate. *Med Drug Discov* 9:100077. <https://doi.org/10.1016/j.medidd.2020.100077>
25. Aghdam R, Habibi M, Taheri G (2021) Using informative features in machine learning based method for COVID-19 drug repurposing. *J Cheminform* 13(1):1–14. <https://doi.org/10.1186/s13321-021-00553-9>
26. Jamshidi MB, Talla J, Lalbakhsh A, Sharifi-Atashgah MS, Sabet A, Peroutka Z (2021) A conceptual deep learning framework for COVID-19 drug discovery. In: 2021 IEEE 12th annual ubiquitous computing, electronics & mobile communication conference (UEMCON), IEEE. <https://doi.org/10.1109/uemcon53757.2021.9666715>
27. Sharma R, Shrivastava S, Singh SK, Kumar A, Singh AK, Saxena S (2021) Deep-AVPpred: artificial intelligence driven discovery of peptide drugs for viral infections. *IEEE J Biomed Health Inform* 26:1–1. <https://doi.org/10.1109/jbhi.2021.3130825>
28. Ramadhanti NS, Kusuma WA, Batubara I, Heryanto R (2021) Random forest to predict eucalyptus as a potential herb in preventing covid19. In: 2021 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), IEEE. <https://doi.org/10.1109/cibcb49929.2021.9562940>
29. Chou H-C, Cheng C-H, Chang J-R (2007) Extracting drug utilization knowledge using self-organizing map and rough set theory. *Expert Syst Appl* 33(2):499–508. <https://doi.org/10.1016/j.eswa.2006.05.020>
30. Sujuan Z, Tao Z, Yongquan Y (2009) The extension-based ant colony optimization of finding rough set reducts. *International forum on information technology and applications*. IEEE, Piscataway
31. Wang X, Yang J, Teng X, Xia W, Jensen R (2007) Feature selection based on rough sets and particle swarm optimization. *Pattern Recognit Lett* 28(4):459–471
32. Alazzam H, Sharieh A, Sabri KE (2020) A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. *Expert Syst Appl* 148:113249. <https://doi.org/10.1016/j.eswa.2020.113249>
33. Gupta A, Purohit A (2017) RGAP: a rough set, genetic algorithm and particle swarm optimization based feature selection approach. *Int J Comput Appl* 161(6):1–5. <https://doi.org/10.5120/ijca2017913228>
34. Pawlak Z (1991) *Rough sets*. Springer, Cham
35. Chen Y, Zhu Q, Xu H (2015) Finding rough set reducts with fish swarm algorithm. *Knowl Based Syst* 81:22–29. <https://doi.org/10.1016/j.knsys.2015.02.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ibrahim Gad¹  · Mohamed Torky² · Yaseen A. M. M. Elshaier^{3,7} · Ashraf Darwish^{4,7} · Aboul Ella Hassanien^{5,6,7}

✉ Ibrahim Gad
ibrahim.gad@science.tanta.edu.eg

Mohamed Torky
mohamed-turki@eru.edu.eg

Yaseen A. M. M. Elshaier
Yaseen.elshaier@fop.usc.edu.eg

Ashraf Darwish
ashraf.darwish.eg@ieee.org

Aboul Ella Hassanien
abo.aly@ku.edu.kw; aboitcairo@cu.edu.eg;
<http://www.egyptscience-srge.com>

¹ Department of Computer Science, Faculty of Science, Tanta University, Tanta, Egypt

² Faculty of Artificial Intelligence, Egyptian Russian University, Cairo, Badr City 11829, Egypt

³ Organic and Medicinal Chemistry Department, Faculty of Pharmacy, University of Sadat City, Sadat City 32897, Egypt

⁴ Faculty of Science, Helwan University, Helwan, Cairo, Egypt

⁵ College of Business Administration, Kuwait University, Kuwait, Kuwait

⁶ Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt

⁷ Scientific Research School of Egypt (SRSEG), Cairo, Egypt