*Research Article*

# Discovery and Validation of Traditional Chinese and Western Medicine Combination Antirheumatoid Arthritis Drugs Based on Machine Learning (Random Forest Model)

**Jijia Sun** [ID],[1,2] **Qinghua Ni,**[2] **Fengyan Jiang,**[1] **Baocheng Liu,**[1] **Jianying Wang** [ID],[1] **Lei Zhang** [ID],[1] **and Jihan Huang** [ID][3]

[1]*Shanghai Collaborative Innovation Center of Traditional Chinese Medicine Health Service, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[2]*School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[3]*Center for Drug Clinical Research, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

Correspondence should be addressed to Jianying Wang; wjy8310@163.com, Lei Zhang; zhanglei37@sina.com, and Jihan Huang; huangjihan@shutcm.edu.cn

The combination of traditional Chinese medicine (TCM) and Western medicine is a promising method for treating rheumatoid arthritis (RA). Combining the two fully exploits the advantages of Western and TCM to treat RA and has the potential to greatly improve the therapeutic effect on RA. In this study, we developed a combination drug training set by using 16 characteristic variables based on the characteristics of small molecules of TCM ingredients and Food and Drug Administration-certified combination drug data downloaded from the DrugCombDB database. Furthermore, we compared the prediction and classification abilities of five models: the $k$-nearest neighbors, naive Bayes, support vector machine, random forest, and AdaBoost algorithms. The random forest model was selected as the classification and prediction model for Western and TCM and Western combination drugs. We collected data for 41 small molecules of TCM ingredients from the Traditional Chinese Medicine Systems Pharmacology database and 10 small molecule drugs commonly used in anti-RA treatment from the DrugBank database. Combinations of Western and TCM for anti-RA treatment were screened. Finally, the CellTiter-Glo method was used to determine the synergy of these combinations, and the 15 most predicted drug combinations were carried out experimental verification. Myricetin, rhein, nobiletin, and fisetin had high synergy with celecoxib, and rhein had high synergy with hydroxychloroquine. The preliminary findings of this study can be further applied for practical clinical anti-RA combined treatment strategies and serve as a reference for clinical treatment of RA with integrated Western and TCM.

## 1. Introduction

Modern medical research has demonstrated that many diseases often present complex pathogenic mechanisms, and consequently, the past methods for treating diseases with drugs are no longer adequate. In recent years, drug combination has been reported to offer clear advantages in treating tumors, AIDS, cardiovascular diseases, osteoarthritis, and other complex diseases; it has been widely implemented in clinical practice [1–4], with treatment methods based on combined drugs demonstrating advantageous synergistic effects, or a "$1 + 1 \geq 2$" effect. The therapeutic effects of two or more drug combinations are greater than those achieved when each drug is used alone. Therefore, developing and evaluating potentially synergetic combination drugs, whether for development of new drugs or clinical treatment optimization, are essential. However, most existing combination drugs have been developed on the basis of limited clinical experience or experimental strategies. Verified combination drugs with synergistic therapeutic effects currently

comprise a small portion of listed drugs. Therefore, making full use of knowledge of existing drug combinations, discovering new drug combinations based on well-accepted and understood drugs, and establishing a model for evaluating the synergistic effects of drug combinations are key concerns that must be addressed.

Rheumatoid arthritis (RA) is a chronic, systemic, inflammatory disease of the synovial and other joints. RA is a chronic progressive autoimmune disease [5], and its underlying pathogenesis is currently unclear. However, studies have reported that the incidence of RA increases with age due to infection and inflammatory mediators, and approximately 0.3%–1.0% of the global population is affected by the disease every year [6, 7]. At present, the drugs used in clinical treatment of RA are mainly nonsteroidal anti-inflammatory drugs (NSAIDs), disease modifying antirheumatic drugs (DMARDs), and biological macromolecular therapy. Because RA is complex, several studies have explored different combination therapy strategies for its treatment [8–10]. However, some therapeutic drugs, such as methotrexate, may produce unfavorable side effects [11, 12]. Therefore, if sole use of anti-RA combination therapy has the potential to produce unfavorable side effects, long-term use in patients with RA may do greater harm. Traditional Chinese medicine (TCM) has long been used to treat RA, and several studies have found that combined anti-RA TCM and a compound prescription have not only anti-inflammatory, analgesic, immune regulation, and multilevel and multilink therapeutic effects but also the advantages of high safety, few adverse reactions, and low cost. Such a combination has gained attention in studies of RA treatment and has increasingly attracted international attention [13, 14].

Therefore, we propose that the combination of Western medicine and TCM may be a promising method for improving the efficacy of anti-RA treatments and reducing side effects. Through their combination, the advantages of both Western medicine and TCM can be fully exploited to improve the therapeutic effects of RA treatment. However, because numerous active ingredients in TCM and hundreds of listed Western medicines have been reported to be effective in anti-RA treatments, the potential drug combinations are in the thousands or even tens of thousands. This multitude of combinations cannot feasibly be evaluated or verified through clinical practice and trials; therefore, the most practical method for identifying potential combinations would be to use drug data for existing combination drugs to model, evaluate, and screen potential anti-RA combinations of Western medicine and TCM and then experimentally validate these models. Therefore, in this study, we developed a classification and prediction model for combining Western and TCM drugs based on characteristics obtained through small molecule research on TCM and existing data on combined drugs. This model was developed using machine learning modeling followed by manual screening of anti-RA combination drugs. Through experimental verification, we identified effective combinations of Western and TCM drugs for treating RA; these findings can serve as a reference for clinical treatment of RA by using such combined treatment. The flow chart illustrating the process of drug discovery and verification for anti-RA integrative medicine through machine learning modeling is displayed in Figure 1.

## 2. Materials and Methods

### 2.1. Data Collection and Collation.
The DrugCombDB database (http://drugcombdb.denglab.org/) is a comprehensive combinatorial drug database [15] containing integrated drug combinations from various data sources. To ensure the reliability of the research data, we first downloaded the Food and Drug Administration- (FDA-) provided known drug combination dataset from the database (up to May 31, 2019). The molecular formula, Chemical Abstracts Service (CAS) number, canonical simplified molecular-input line-entry system (SMILES), and other chemical informatics data corresponding to each combination of drugs and their action targets were collected from the DrugBank database (https://go.drugbank.com/).

Furthermore, we searched "Rheumatoid Arthritis" on the Traditional Chinese Medicine Systems Pharmacology (TCMSP) database (https://tcmspw.com/) to retrieve target Chinese medicine ingredient data reported to be relevant to treating RA. We obtained the main small molecules chemical information from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). In addition, several common anti-RA drugs were selected from the DrugBank database along with those discussed in the literature and those recommended by clinicians. All small molecule data for the TCM ingredients (collected from the TCMSP) and for known anti-RA Western drugs were combined in pairs to form the final prediction sample set.

### 2.2. Construction of Combined Drug Characteristic Variables

#### 2.2.1. Drug Combination Features Based on Molecular Fingerprint Similarity.
Canonical SMILES of all drugs were collected from the DrugBank and PubChem databases, and the 1024-dimensional molecular fingerprints for each of the small molecule drugs were calculated using DRAGON software (version 7.0). The fingerprint similarity ($S_{FP}$) between two small molecule drugs was then calculated using Tanimoto coefficient-based similarity:

$$S_{FP}\langle d_A, d_B \rangle = \frac{A \bullet B}{\|A\|^2 + \|B\|^2 - A \bullet B}, \tag{1}$$

where $A$ is the 1024-dimensional molecular fingerprint feature vector corresponding to the small molecule drug $d_A$, $\|A\|$ is the vector length, $B$ is the 1024-dimensional molecular fingerprint feature vector corresponding to the small molecule drug $d_B$, $\|B\|$ is the vector length, and $A \bullet B$ is the vector inner product.

#### 2.2.2. Drug Combination Characteristics Based on Small Molecule ADME Similarity.
To predict the adsorption, distribution, metabolism, and excretion (ADME) characteristics of each drug, Canonical SMILES for all small molecule drugs were imported into the SwissADME online system (http://www.swissadme.ch/) [16]. We selected 12 indicators as characteristic indexes for our ADME similarity calculations:

molecular weight, heavy atoms, aromatic heavy atoms, fraction Csp3, rotatable bonds, H-bond acceptors, H-bond donors, molecular refractivity, topological polar surface area, Silicos-IT Log P, and Silicos-IT LogSw. All ADME indexes were first normalized, and the Euclidean distance was then used to calculate the similarity value $S_{\text{ADME}}$ of ADME:

$$S_{\text{ADME}} = \left[ \sum_{i=1}^{n} \left( x_i^A - x_i^B \right)^2 \right]^{1/2}, \quad (2)$$

where $x_i^A$ is the ADME index corresponding to the drug small molecule $d_A$ and $x_i^B$ is the ADME index corresponding to the drug small molecule $d_B$ ($n = 12$).

*2.2.3. Drug Combination Characteristics Based on Sequence Similarity of Small Molecular Targets.* The action targets of all drugs were obtained from DrugBank and the TCMSP, and the protr package (version 1.6.2) based on R language was used to obtain the sequence of action protein targets of all small molecule drugs. The sequence similarity between the two target proteins, $S_{\text{Seq}}\langle d_A, d_B \rangle$, was calculated using the Smith–Waterman algorithm. The maximum, minimum, median, and mean values of the sequence similarity $S_{\text{Seq}}\langle d_A, d_B \rangle$ between all target proteins of the $d_A$ and $d_B$ were included in the characteristic variables as $S_{\text{Seqmax}}$, $S_{\text{Seqmin}}$, $S_{\text{Seqmed}}$, and $S_{\text{Seqmean}}$, respectively.

*2.2.4. Drug Combination Characteristics Based on GO Functional Similarity of Small Molecular Targets.* To calculate the gene ontology (GO) functional similarity $S_{\text{GO}}\langle d_A, d_B \rangle$ of the two drugs $d_A$ and $d_B$, we used the GOSemSim package (version 2.14.2). Similarly, the minimum, maximum, median, and mean values of the GO functional similarity $S_{\text{GO}}\langle d_A, d_B \rangle$ of the two drugs $d_A$ and $d_B$ were included in the characteristic variables as $S_{\text{GOmin}}$, $S_{\text{GOmax}}$, $S_{\text{GOmed}}$, and $S_{\text{GOmean}}$, respectively.

*2.2.5. Drug Combination Characteristics Based on Pathway Similarity of Small Molecule Targets.* We used the cluster-Profiler package (version 3.14.3) to perform a Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway enrichment analysis on all drug small molecule targets. We then screened the enrichment pathways of all targets based on $p$ value < 0.05 and $q$ value < 0.05. A vector, $P = (p_1, p_2, \cdots, p_n)$, consisting of all pathways was constructed according to the enriched pathways, $\{p_k | k = 1, 2, \cdots, n\}$ (where $n$ is the number of pathways). If a target was enriched in a pathway $p_k$ in the target set $T_A$ of the drug small molecule $d_A$, then the value was set as $p_k^A = 1$ at the corresponding position of the drug small molecule $A$ pathway vector $P_A$, and if not, it was set as 0:

$$p_k^A = \begin{cases} 1, & T_A \cap p_k \neq \varnothing, \\ 0, & \text{others.} \end{cases} \quad (3)$$

The target pathway vector $P_B$ of the drug small molecule $d_B$ can be similarly obtained using a string similarity formula to calculate the pathway similarity value ($S_{\text{Pathway}}$) of two drugs:

$$S_{\text{Pathway}}\langle d_A, d_B \rangle = \frac{P_A \bullet P_B}{\|P_A\| \times \|P_B\|}, \quad (4)$$

where $\|P_A\|$ and $\|P_B\|$ are path vector lengths and $P_A \cdot P_B$ is the vector inner product.

*2.2.6. Drug Combination Characteristics Based on Distance of Small Molecule Targets in Human PPI Network.* The distance and network proximity of small molecule targets of different combinations of drugs on the human protein–protein interaction (PPI) network were calculated to measure the synergy characteristics of combination drugs within the PPI network:

$$S_{\text{PPImin}}\langle d_A, d_B \rangle = \frac{1}{1 + e^{-\min (d(a,b))}},$$

$$S_{\text{PPImax}}\langle d_A, d_B \rangle = \frac{1}{1 + e^{-\max (d(a,b))}},$$

$$S_{\text{PPImed}}\langle d_A, d_B \rangle = \frac{1}{1 + e^{-\text{med}(d(a,b))}},$$

$$S_{\text{PPImean}}\langle d_A, d_B \rangle = \frac{1}{1 + e^{-\langle d_{AB} \rangle}},$$

$$S_{\text{np}}\langle d_A, d_B \rangle = S_{\text{PPImean}}\langle d_A, d_B \rangle - \frac{S_{\text{PPImean}}\langle d_A, d_A \rangle + S_{\text{PPImean}}\langle d_B, d_B \rangle}{2}, \quad (5)$$

where $d(a, b)$ is the shortest path distance between two protein targets, $a$ and $b$, within the PPI network (calculated using the igraph package [version 1.2.6]), $\langle d_{AB} \rangle = 1/\|A\| \times \|B\| \sum_{a \in A, b \in B} d(a, b)$, $\|A\|$ is the number of targets for the drug small molecule $d_A$, and $\|B\|$ is the number of targets for the drug small molecule $d_B$. The human PPI background network used in the calculation was derived from the literature [17]; the PPI network contained a total of 16677 proteins and 243603 interactions.

The Tanimoto coefficient-based similarity, Euclidean distance, and cosine similarity calculations used in this paper were calculated using the philentropy package (version 0.4.0) based on R language (version 4.0.2).

*2.3. Construction and Selection of Combination Drug Prediction Model.* To comprehensively determine and select an appropriate machine learning model for modeling and prediction in this study, we selected different machine learning classifiers for use with the training set and the verification set and compared their classification accuracies. In this study, we selected and compared five classical machine learning models: the $k$-nearest neighbors ($k$-NN) [18], naive Bayes (NB) [19], support vector machine (SVM) [20], random forest (RF) [21], and AdaBoost classifiers [22]. The caret package (version 6.0-86), the klaR package (version 0.6-15), the svmRadial model in the kernlab package (version 0.9-29), the randomForest package (version 4.6-14), and the ada package (version 2.0-5) were used for the $k$-NN, naive Bayes, support vector machine, random forest, and AdaBoost classifiers, respectively.
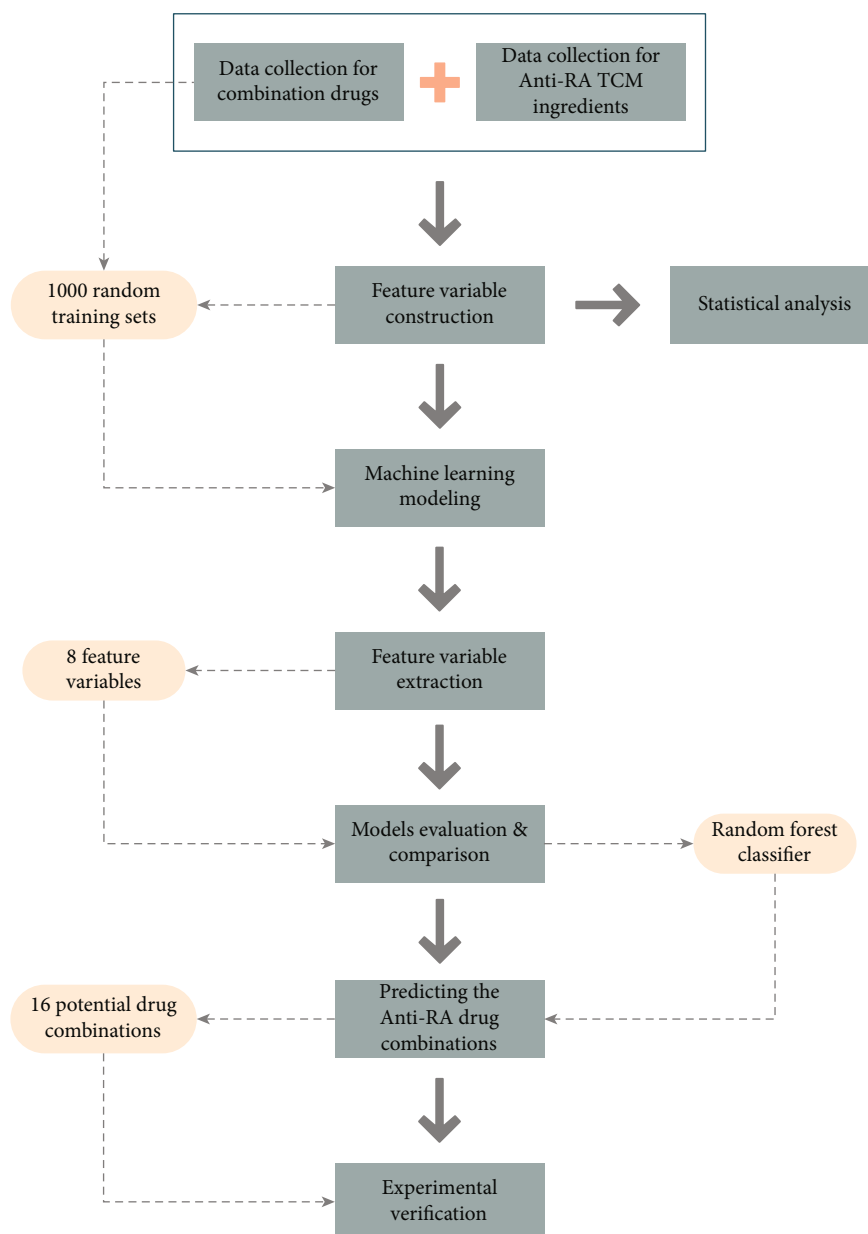
FIGURE 1: Flow chart of drug discovery and verification for anti-RA integrative medicine through machine learning modeling.

Furthermore, we used established sample training sets to compare the classification prediction effects of the five machine learning models. First, the sample set was divided into prediction training sets and internal validation sets at a 7 : 3 ratio. Then, the prediction training set was input into the five machine learning models, and the machine learning models were trained using three sets of 10-fold cross-validation. The trained machine learning models were evaluated using the internal validation sets, and the model with the best classification prediction effect was selected as the final drug prediction model for the Western and TCM drug combinations.

2.4. Evaluation of Machine Learning Models. In this study, we defined precision rate as the proportion of positive samples correctly predicted by the model out of all positive predictions, recall rate as the proportion of positive samples correctly predicted by the model out of all predictions, accuracy as the proportion of positive and negative samples to all samples correctly predicted by the model, and $F1$-score as the overall score of the model when precision rate and recall rate were given equal weight (calculated using Formulas (6)–(9), respectively). A higher score indicated a better classification effect.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (6)$$

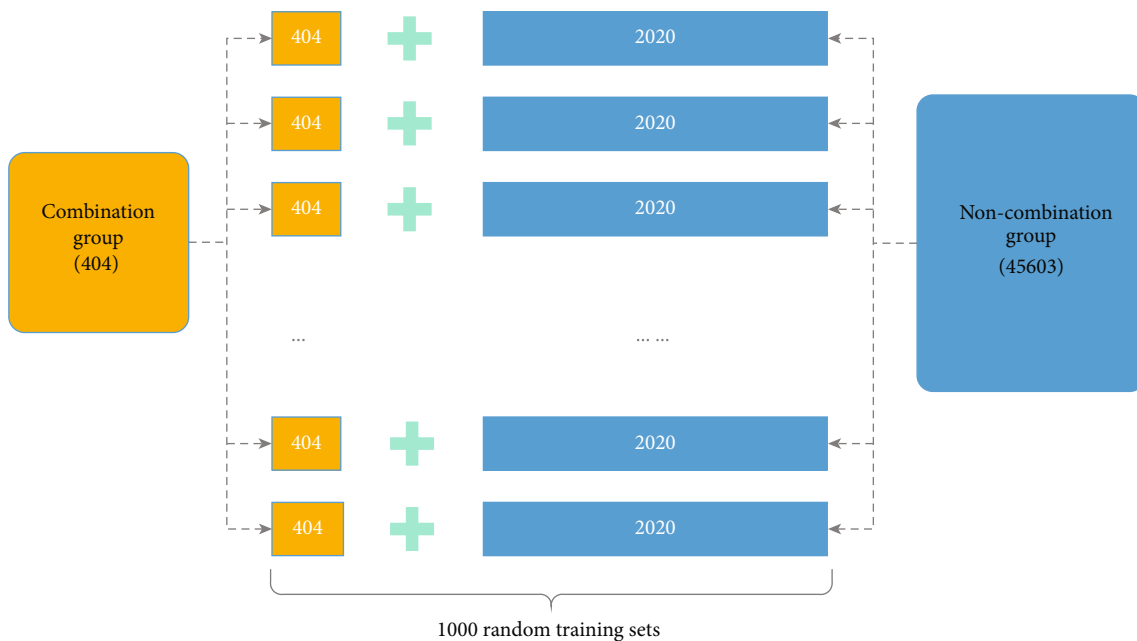$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (7)$$

FIGURE 2: Training set of prediction samples comprised 1000 training sets of prediction samples randomly selected from 2020 samples of noncombination drugs and 404 samples of combination drugs.

TABLE 1: Results of $t$-test and Mann–Whitney nonparametric test for 16 characteristic variables between different groups.

| Feature | Noncombination group | Combination group | $t$-test | Mann–Whitney test |
|---|---|---|---|---|
| $S_{FP}$ | $0.783 \pm 0.087$ | $0.783 \pm 0.091$ | 0.914 | 0.664 |
| $S_{ADME}$ | $3.347 \pm 2.994$ | $3.160 \pm 2.697$ | 0.167 | 0.199 |
| $S_{Seqmin}$ | $0.014 \pm 0.045$ | $0.020 \pm 0.070$ | 0.109 | 0.400 |
| $S_{Seqmax}$ | $0.066 \pm 0.189$ | $0.173 \pm 0.342$ | ** | ** |
| $S_{Seqmed}$ | $0.021 \pm 0.057$ | $0.030 \pm 0.084$ | * | ** |
| $S_{Seqmean}$ | $0.026 \pm 0.062$ | $0.049 \pm 0.105$ | ** | ** |
| $S_{GOmin}$ | $0.236 \pm 0.175$ | $0.221 \pm 0.193$ | 0.114 | ** |
| $S_{GOmax}$ | $0.523 \pm 0.239$ | $0.550 \pm 0.288$ | 0.061 | 0.092 |
| $S_{GOmed}$ | $0.374 \pm 0.172$ | $0.369 \pm 0.205$ | 0.649 | 0.532 |
| $S_{GOmean}$ | $0.376 \pm 0.166$ | $0.374 \pm 0.194$ | 0.825 | 0.644 |
| $S_{Pathway}$ | $0.007 \pm 0.029$ | $0.014 \pm 0.039$ | ** | ** |
| $S_{PPImin}$ | $0.867 \pm 0.092$ | $0.821 \pm 0.142$ | ** | ** |
| $S_{PPImax}$ | $0.931 \pm 0.116$ | $0.934 \pm 0.094$ | 0.500 | 0.889 |
| $S_{PPImed}$ | $0.917 \pm 0.043$ | $0.910 \pm 0.050$ | ** | ** |
| $S_{PPImean}$ | $0.100 \pm 0.064$ | $0.121 \pm 0.091$ | ** | ** |
| $S_{np}$ | $-0.487 \pm 0.215$ | $-0.436 \pm 0.216$ | ** | ** |

$*p < 0.05$ and $**p < 0.01$.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \qquad (8)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \qquad (9)$$

where true positive (TP) indicates a sample is positive and was predicted to be positive, false negative (FN) indicates a sample is positive but was predicted to be negative, false positive (FP) indicates a sample is negative but was predicted to be positive, and true negative (TN) indicates a sample is negative and was predicted to be negative.

In the plot of the receiver operating characteristic (ROC) curve, the horizontal axis represents the false positive rate (FPR) and the vertical axis represents the true positive rate (TPR). The area under the ROC curve is the area surrounded by the ROC and its lower coordinate axis and is represented as AUROC in this paper. In addition, because the precision and recall rates of the model are not fixed, when the classification threshold of the model is adjusted, the precision and recall rates continue to change. The precision–recall curve represents the relationship between the precision rate of the ordinate and the recall rate of the abscissa with the change of the classification threshold $p$ value; the area beneath this curve is represented as the area under the precision–recall curve (AUPRC).

### 2.5. In Vitro Cell Experiment Verification

*2.5.1. Cell Culture and Reagents.* RAW264.7 cells were purchased from the American Type Culture Collection. The culture reagents included Dulbecco's modified Eagle medium (DMEM; Gibco, C11995500CP), Roswell Park Memorial Institute (RPMI) 1640 (Gibco, C11875500BT), fetal bovine serum (Bio IND, 04-002-1A), antimycotic (Lifetechnologies, 15240-112), phosphate-buffered saline (PBS), pH 7.4 (Gibco, 10010-023), trypsin-EDTA (0.05%; Lifetechnologies, 25300-054), bovine serum albumin (Lifetechnologies, 15561012), and CellTiter-Glo (CTG; PROMEGA, G7572). Consumables and instruments included a cell culture plate (Corning), cell
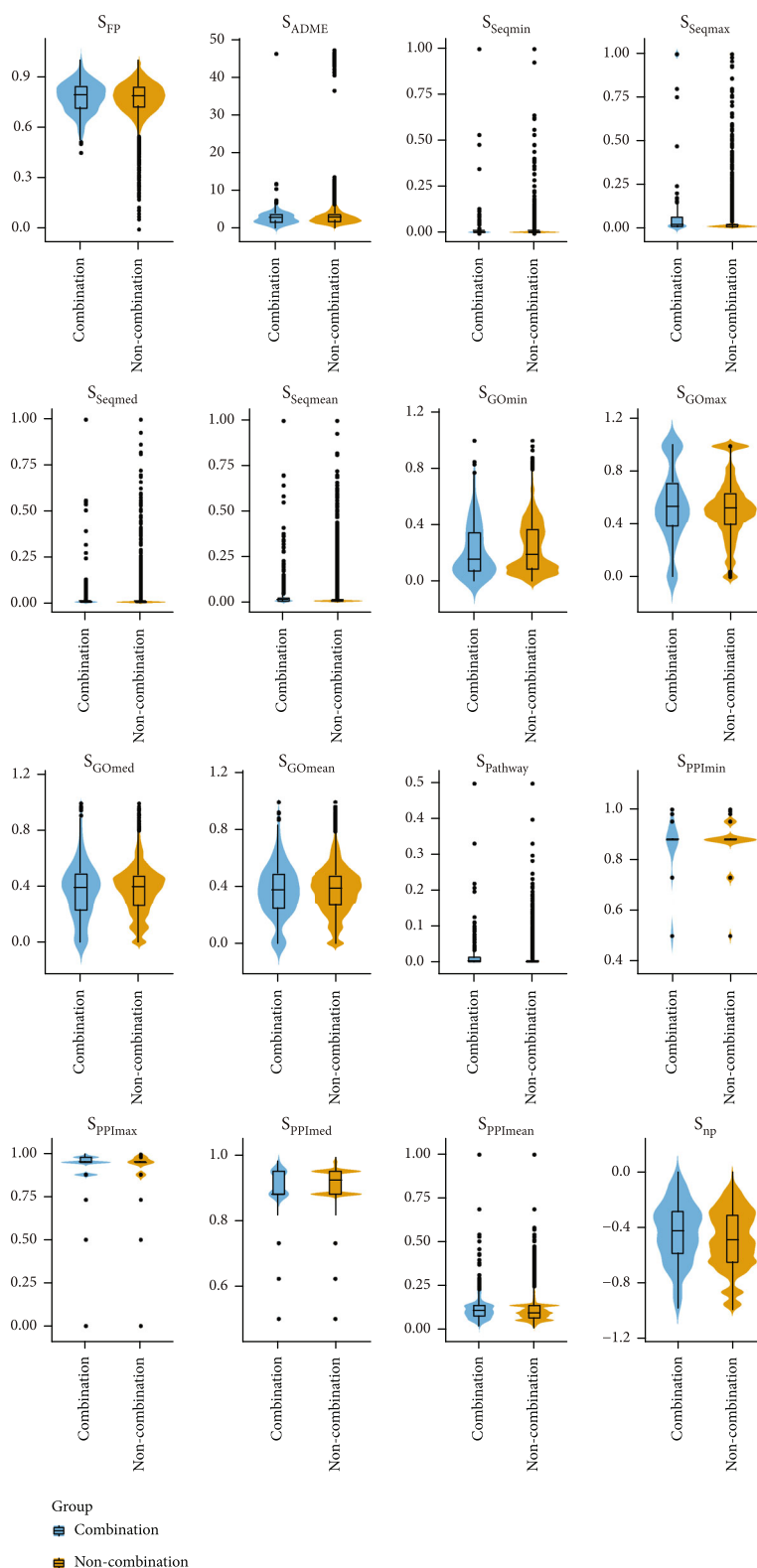
FIGURE 3: Violin diagram of value distribution of 16 characteristic variables in training sample set.

culture flask (Corning), microplate tester (BioTek, HM-1), and conventional instruments, such as a $CO_2$ incubator (Thermo 3111) and biosafety cabinet (Heal Force, HFSAFe-1200LC).

2.5.2. Experimental Procedure and Methods. First, the different compounds were diluted with DMEM medium to the desired concentrations of the experimental design, namely, 0, 0.032, 0.16, 0.8, 4, 20, and 100 $\mu$M. The cells were then
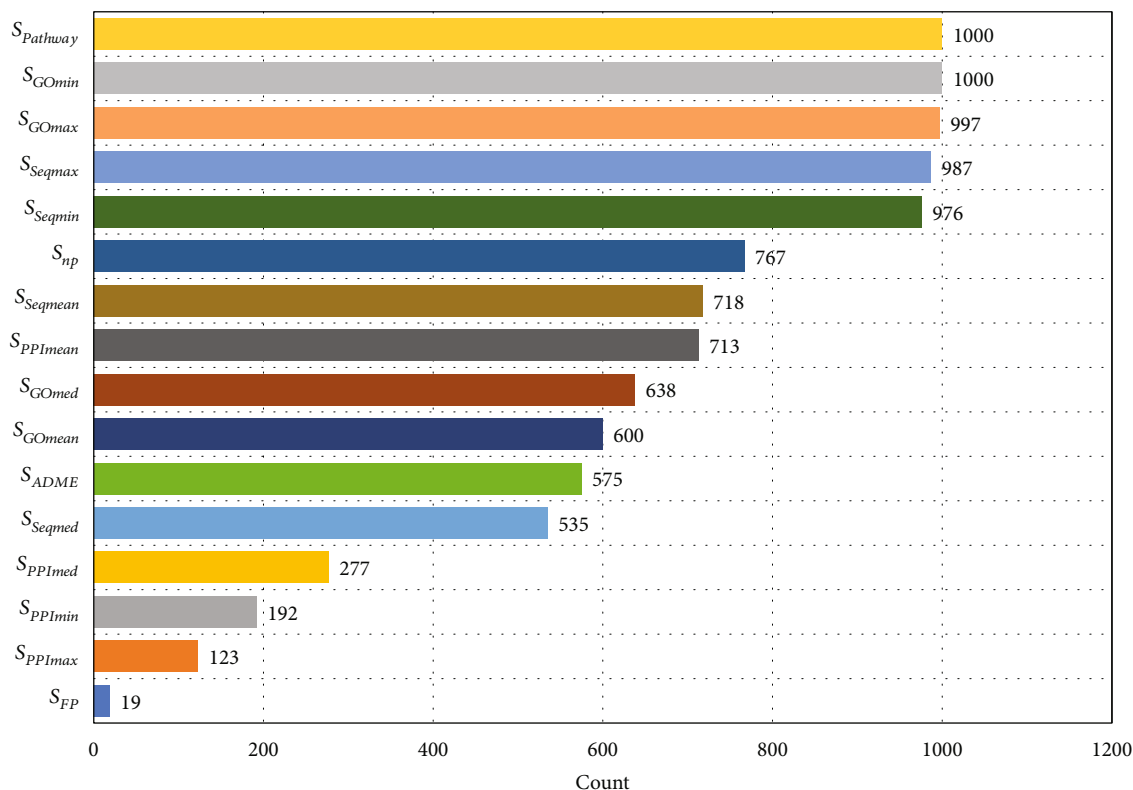
FIGURE 4: Occurrence of 16 feature variables from 1000 feature extractions.

counted and inoculated into 96-well culture plates at 2000 cells/100 μL/well; 90 μL of culture medium (including serum) was added to each well for overnight culture, and 10 μL of drug solution was added. After a 48 h culture, CTG was used to detect the proliferation of RAW264.7 cells; CTG (100 μL) was added to each well, the content of which was incubated at room temperature in a dark environment for 10 min. The chemiluminescence value of each well was measured at 500 ms using enzyme calibration.

*2.5.3. Combination Index Measurement.* In this study, according to the principle of the median effect—specifically by employing the Chou–Talalay model [1]—the synergistic effect between each set of drugs was analyzed, and the effects of integrated Western and TCM drugs on RA cell lines were evaluated using the combination index (CI). The CI indicates the degree of drug synergy, addition, or antagonism for any given drug combination:

$$\text{CI} = \frac{D_A}{(D_A)_x} + \frac{D_B}{(D_B)_x}, \tag{10}$$

where when $\text{CI} = 1$, the two drugs have an additive effect, when $\text{CI} < 1$, the two drugs have a synergistic effect, and when $\text{CI} > 1$, the two drugs are antagonistic. $D_A$ and $D_B$ represent the concentrations of the two drugs, $d_A$ and $d_B$, when administered in combination, and $(D_A)_x$ and $(D_B)_x$ represent the concentrations of the two drugs, $d_A$ and $d_B$, when they have been administered alone and the combined drug

inhibition rate has been reached.

$$D_x = D_m \left(\frac{f_a}{f_u}\right)^{1/m}. \tag{11}$$

In Formula (11), $f_a$ and $f_u$ are the inhibition and survival rates, respectively, when the two drugs are combined, $f_a = 1 - f_u$, $D_m$ is the median dose, and $m$ is the curve coefficient (calculated according to half-maximal inhibitory concentration ($IC_{50}$) theory).

In this study, CompuSyn software (http://www.combosyn.com/) was used to calculate the CI values of the drug combinations to evaluate whether each screened combination of Western and TCM drugs had synergistic therapeutic effects.

*2.6. Statistical Analysis.* Data were expressed as the mean ± standard deviation, and a *t*-test and Mann–Whitney nonparametric tests were performed using IBM SPSS Statistics (version 26.0).

# 3. Results

*3.1. Collection and Collation of Combined Drug Data.* We obtained 946 pairs (comprising 816 drugs) of combinatorial drug data reported by the FDA from the DrugCombDB database. We further downloaded all existing small molecule drug data from the DrugBank database. After name matching and sorting, some small molecule drug data without key information, such as targets, were removed. In total,

Table 2: Prediction evaluation results of five machine learning classification models under 1000 random sample training sets.

| Model | Accuracy | Precision | Recall | F1 score | AUROC | AUPR |
| --- | --- | --- | --- | --- | --- | --- |
| $k$-NN | $0.827 \pm 0.012$ | $0.848 \pm 0.012$ | $0.965 \pm 0.013$ | $0.903 \pm 0.008$ | $0.651 \pm 0.024$ | $0.309 \pm 0.033$ |
| NB | $0.813 \pm 0.014$ | $0.845 \pm 0.012$ | $0.949 \pm 0.013$ | $0.894 \pm 0.009$ | $0.617 \pm 0.026$ | $0.269 \pm 0.031$ |
| SVM | $0.834 \pm 0.011$ | $0.838 \pm 0.012$ | $0.992 \pm 0.006$ | $0.909 \pm 0.007$ | $0.657 \pm 0.027$ | $0.330 \pm 0.038$ |
| RF | $0.849 \pm 0.012$ | $0.861 \pm 0.012$ | $0.977 \pm 0.008$ | $0.915 \pm 0.007$ | $0.739 \pm 0.024$ | $0.460 \pm 0.039$ |
| AdaBoost | $0.833 \pm 0.011$ | $0.834 \pm 0.012$ | $0.998 \pm 0.004$ | $0.909 \pm 0.007$ | $0.667 \pm 0.027$ | $0.317 \pm 0.038$ |



(a)                                                                                                  (b)
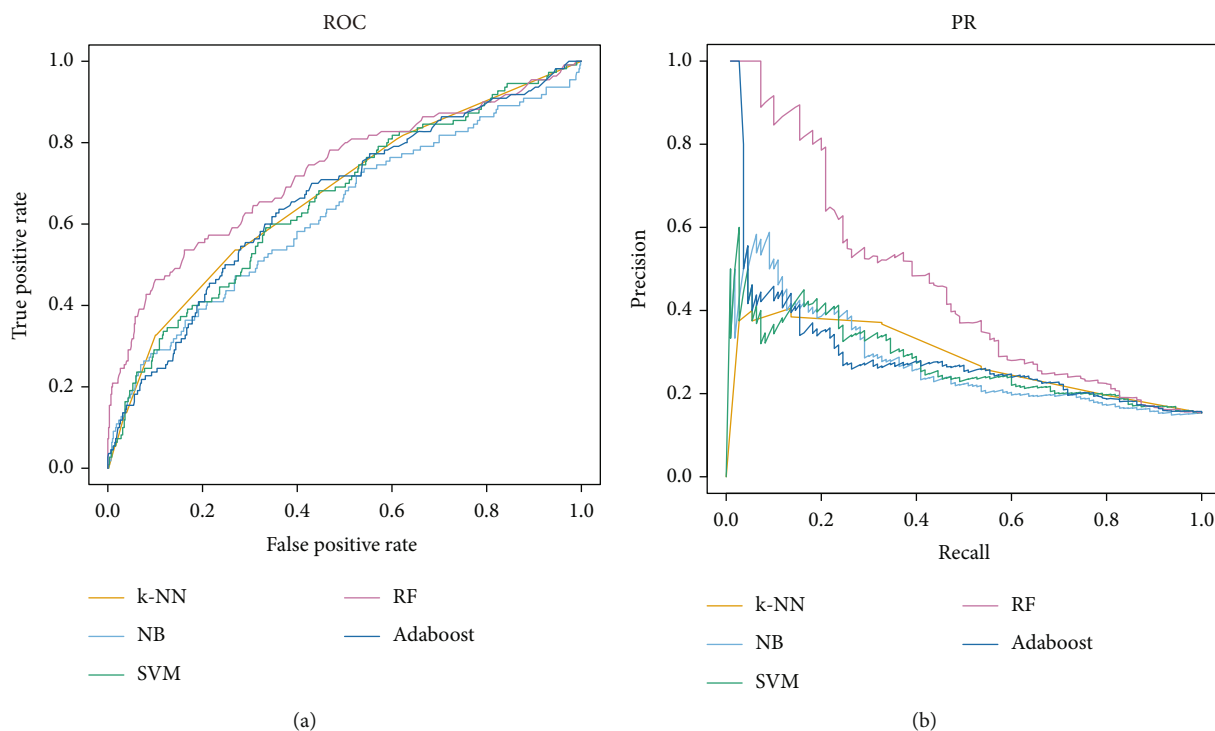
Figure 5: ROC and PR curves of five machine learning models for classification prediction.

we included 488 drugs and their corresponding 546 pairs of combinations. The data obtained for each small molecule drug mainly included DrugBank ID, canonical SMILES, and known targets (Supplementary 1; pretreated FDA-certified combination drug dataset).

*3.2. Construction of Combination Drug Training Set.* First, using the method described in Section 2.2, we obtained the molecular fingerprint similarity value ($S_{FP}$), the ADME index similarity value ($S_{ADME}$), and the minimum, maximum, median, and mean of the similarity ($S_{Seqmin}$, $S_{Seqmax}$, $S_{Seqmed}$, and $S_{Seqmean}$, respectively) of the drug target sequence between the two small molecule drugs. The minimum, maximum, median, and mean values of the GO functional similarity of the drug targets were $S_{GOmin}$, $S_{GOmax}$, $S_{GOmed}$, and $S_{GOmean}$, respectively. The KEGG pathway similarity value ($S_{Pathway}$); the minimum, maximum, median, and mean of the shortest path of different drug targets on the human PPI network ($S$ PPImin, $S_{PPImax}$, $S_{PPImed}$, and $S_{PPImean}$, respectively); and the proximity ($S_{np}$) of the different drug targets were also obtained. Thus, we obtained a total of

16 characteristic variables to predict combinations of Western and TCM drugs.

In this study, we evaluated 16 characteristic variables between two drugs for 546 pairs of combined drugs and removed those for which characteristic variables could not be calculated or were missing. We obtained a total of 404 positive samples of combined drugs (comprising 488 drugs) reported by the FDA. Due to a lack of effective noncombination drug data, we combined the 488 drugs in pairs, removed the existing combination drug samples and those missing data, and obtained a total of 45603 noncombination drugs as negative samples (Supplementary 2; positive and negative sample training dataset for all drug combinations).

Finally, to construct the training set for the TCM-combined drug samples, because the number of noncombined drugs should be much larger than that of combined drugs and to avoid large deviation in positive and negative samples, we set the ratio of positive to negative samples as $1 : 5$; the number of positive samples was 404, and the number of negative samples randomly extracted from the 45603 noncombined drug samples was $404 \times 5 = 2020$. Through

TABLE 3: Small molecules of 41 anti-RA TCM ingredients collected from TCMSP database.

| TCMSP ID | Name | Molecular formula | PubChem CID | CAS |
|---|---|---|---|---|
| MOL006505 | (-)-Epicatechin | C15H14O6 | 72276 | 490-46-0 |
| MOL006821 | (-)-Epigallocatechin gallate | C22H18O11 | 65064 | 989-51-5 |
| MOL008680 | Acetaldehyde | C2H4O | 177 | 75-07-0 |
| MOL000475 | Anethole | C10H12O | 637563 | 104-46-1 |
| MOL000008 | Apigenin | C15H10O5 | 5280443 | 520-36-5 |
| MOL002773 | Beta-carotene | C40H56 | 5280489 | 7235-40-7 |
| MOL000358 | Beta-Sitosterol | C29H50O | 222284 | 83-46-5 |
| MOL003973 | Caffeine | C8H10N4O2 | 2519 | 58-08-2 |
| MOL008842 | Chenodeoxycholic acid | C24H40O4 | 10133 | 474-25-9 |
| MOL000390 | Daidzein | C15H10O4 | 5281708 | 486-66-8 |
| MOL000254 | Eugenol | C10H12O2 | 3314 | 97-53-0 |
| MOL013179 | Fisetin | C15H10O6 | 5281614 | 528-48-3 |
| MOL000392 | Formononetin | C16H12O4 | 5280378 | 485-72-3 |
| MOL000481 | Genistein | C15H10O5 | 5280961 | 446-72-0 |
| MOL002467 | Gingerol | C17H26O4 | 442793 | 23513-14-6 |
| MOL000666 | Hexanal | C6H12O | 6184 | 66-25-1 |
| MOL005916 | Irisolidone | C17H14O6 | 5281781 | 2345-17-7 |
| MOL000422 | Kaempferol | C15H10O6 | 5280863 | 520-18-3 |
| MOL000305 | Lauric acid | C12H24O2 | 3893 | 143-07-7 |
| MOL000006 | Luteolin | C15H10O6 | 5280445 | 491-70-3 |
| MOL007990 | Militarin | C34H46O17 | 171638 | 58139-23-4 |
| MOL002008 | Myricetin | C15H10O8 | 5281672 | 529-44-2 |
| MOL003493 | Naphthalene | C10H8 | 931 | 91-20-3 |
| MOL003403 | Nicotine | C10H14N2 | 89594 | 54-11-5 |
| MOL005828 | Nobiletin | C21H22O8 | 72344 | 478-01-3 |
| MOL006214 | Progesterone | C21H30O2 | 5994 | 57-83-0 |
| MOL012297 | Puerarin | C21H20O9 | 5281807 | 3681-99-0 |
| MOL000098 | Quercetin | C15H10O7 | 5280343 | 117-39-5 |
| MOL012744 | Resveratrol | C14H12O3 | 445154 | 501-36-0 |
| MOL002268 | Rhein | C15H8O6 | 10168 | 478-43-3 |
| MOL011865 | Rosmarinic acid | C18H16O8 | 5281792 | 20283-92-5 |
| MOL006356 | Sorbitol | C6H14O6 | 5780 | 50-70-4 |
| MOL007154 | Tanshinone IIA | C19H18O3 | 164676 | 568-72-9 |
| MOL003186 | Tripterine | C29H38O4 | 122724 | 34157-83-0 |
| MOL003187 | Triptolide | C20H24O6 | 107985 | 38748-32-2 |
| MOL004932 | Uralsaponin A | C42H62O16 | 128229 | 103000-77-7 |
| MOL000511 | Ursolic acid | C30H48O3 | 64945 | 77-52-1 |
| MOL000635 | Vanillin | C8H8O3 | 1183 | 121-33-5 |
| MOL000173 | Wogonin | C16H12O5 | 5281703 | 632-85-9 |
| MOL009357 | Yakuchinone A | C20H24O3 | 133145 | 78954-23-1 |
| MOL009358 | Yakuchinone B | C20H22O3 | 6440365 | 81840-57-5 |

this, a training sample set was formed containing 2424 samples, and the training set was repeated 1000 times to randomly generate 1000 prediction training sets of combined drugs, as displayed in Figure 2.

3.3. Analysis and Screening of Characteristic Variables in Combination Drug Training Set. A $t$-test and a Mann–Whitney nonparametric test were used to determine whether significant differences existed between the 16 characteristic variables in the combined drug group and those in the noncombined drug group ($p < 0.05$). The results revealed that $S_{FP}$, $S_{ADME}$, $S_{Seqmin}$, $S_{GOmax}$, $S_{GOmed}$, $S_{GOmean}$, and $S_{PPImax}$ did not significantly differ between the combined drug and noncombined drug groups ($p > 0.05$). Significant differences

TABLE 4: Information from DrugBank database for 10 common anti-RA small molecule drugs.

| ID | Name | Molecular formula | PubChem CID | CAS |
|---|---|---|---|---|
| DB00316 | Acetaminophen | C8H9NO2 | 1983 | 103-90-2 |
| DB00328 | Indometacin | C19H16ClNO4 | 3715 | 53-86-1 |
| DB00482 | Celecoxib | C17H14F3N3O2S | 2662 | 169590-42-5 |
| DB00563 | Methotrexate | C20H22N8O5 | 126941 | 59-05-2 |
| DB00586 | Diclofenac | C14H11Cl2NO2 | 3033 | 15307-86-5 |
| DB00788 | Naproxen | C14H14O3 | 156391 | 22204-53-1 |
| DB00795 | Sulfasalazine | C18H14N4O5S | 5339 | 599-79-1 |
| DB01097 | Leflunomide | C12H9F3N2O2 | 3899 | 75706-12-6 |
| DB01611 | Hydroxychloroquine | C18H26ClN3O | 3652 | 118-42-3 |
| DB14006 | Choline salicylate | C12H19NO4 | 54686350 | 2016-36-6 |

TABLE 5: Combination data for 15 most predicted drug combinations.

| Group | TCM ingredient | Anti-RA drug | Count |
|---|---|---|---|
| 1 | Eugenol | Acetaminophen | 993 |
| 2 | (-)-Epicatechin | Celecoxib | 957 |
| 3 | Rosmarinic acid | Celecoxib | 946 |
| 4 | Chenodeoxycholic acid | Acetaminophen | 929 |
| 5 | Eugenol | Diclofenac | 912 |
| 6 | Eugenol | Naproxen | 910 |
| 7 | Myricetin | Celecoxib | 908 |
| 8 | Chenodeoxycholic acid | Diclofenac | 898 |
| 9 | Chenodeoxycholic acid | Naproxen | 897 |
| 10 | Rhein | Celecoxib | 894 |
| 11 | Anethole | Naproxen | 854 |
| 12 | Anethole | Diclofenac | 853 |
| 13 | Nobiletin | Celecoxib | 832 |
| 14 | Fisetin | Celecoxib | 791 |
| 15 | Rhein | Hydroxychloroquine | 770 |

were found in the $S_{\text{Seqmax}}$, $S_{\text{Seqmed}}$, $S_{\text{Seqmean}}$, $S_{\text{Pathway}}$, $S_{\text{PPImin}}$, $S_{\text{PPImed}}$, $S_{\text{PPImean}}$, and $S_{\text{np}}$ between the two statistical tests ($p < 0.05$). Statistical differences were found in $S_{\text{GOmin}}$ in the nonparametric tests ($p < 0.05$). All statistical tests were performed in IBM SPSS Statistics (version 26.0), and their data are presented in Table 1. The violin diagrams of the distribution of characteristic variables in different groups were completed using ggplot2 (version 3.3.5) and are displayed in Figure 3.

An *rfe* function using the caret package was used to screen the 16 characteristic variables. A random forest model was used as the selection function, which was *functions = rf Funcs*, and the method was cross-validation, which was *method = "cv."* The others parameters were set to default. The training set was repeated 1000 times to extract feature variables from 1000 randomly generated training set samples, and each extracted feature variable set was recorded. The frequency of each feature variable in the 1000 feature selections was counted; the results are displayed in Figure 4.

As illustrated in Figure 4, eight characteristic variables, namely, $S_{\text{Pathway}}$, $S_{\text{GOmin}}$, $S_{\text{GOmax}}$, $S_{\text{Seqmin}}$, $S_{\text{Seqmax}}$, $S_{\text{Seqmean}}$, $S_{\text{PPImean}}$, and $S_{\text{np}}$, had appearance frequencies of over 70% after 1000 characteristic screenings. $S_{\text{Pathway}}$ and $S_{\text{GOmin}}$ were identified as key characteristic variables for predicting combined drugs. The other eight indicators had appearance frequencies below 70%; for example, $S_{\text{FP}}$ had a rate of less than 10%. Thus, we removed the group of indicators with frequencies lower than 70% and included the top eight indicators as feature variables for subsequent machine learning.

3.4. Comparison and Selection Results of Machine Learning Models. The five included machine learning models were subjected to three replicates of 10-fold cross-validation under 1000 random samples to evaluate the predictive classification performance of different models. The results are displayed in Table 2, and the ROC and PR curves of the five models are presented in Figure 5.

Through evaluation and comparison of the predictive ability of the different models, we found that the classification index scores of the random forest model were generally better than those of the other prediction classifiers. Therefore, the random forest model was adopted as the combined drug prediction classifier for this study. The parameter was represented as *ntree* = 300.

3.5. Prediction Results of Integrative Medicine. We entered "Rheumatoid Arthritis" into the TCMSP database as a search term and identified 41 small molecule types in TCM ingredients that have been used to treat RA. We collected the target treatment information for each small molecule and verified it using the Uniprot database (https://www.uniprot.org/). We collected the molecular formula, PubChem CID, CAS number, and canonical SMILES data of each small molecule from PubChem (Table 3).

The information for drugs commonly used in RA treatment was also obtained from DrugBank. In this study, acetaminophen, celecoxib, choline salicylate, diclofenac, hydroxychloroquine, indometacin, leflunomide, methotrexate, naproxen, sulfasalazine, and 10 other common drugs were selected, as presented in Table 4.

The aforementioned 41 small molecule TCM (collected from TCMSP) and 10 common anti-RA Western medicine
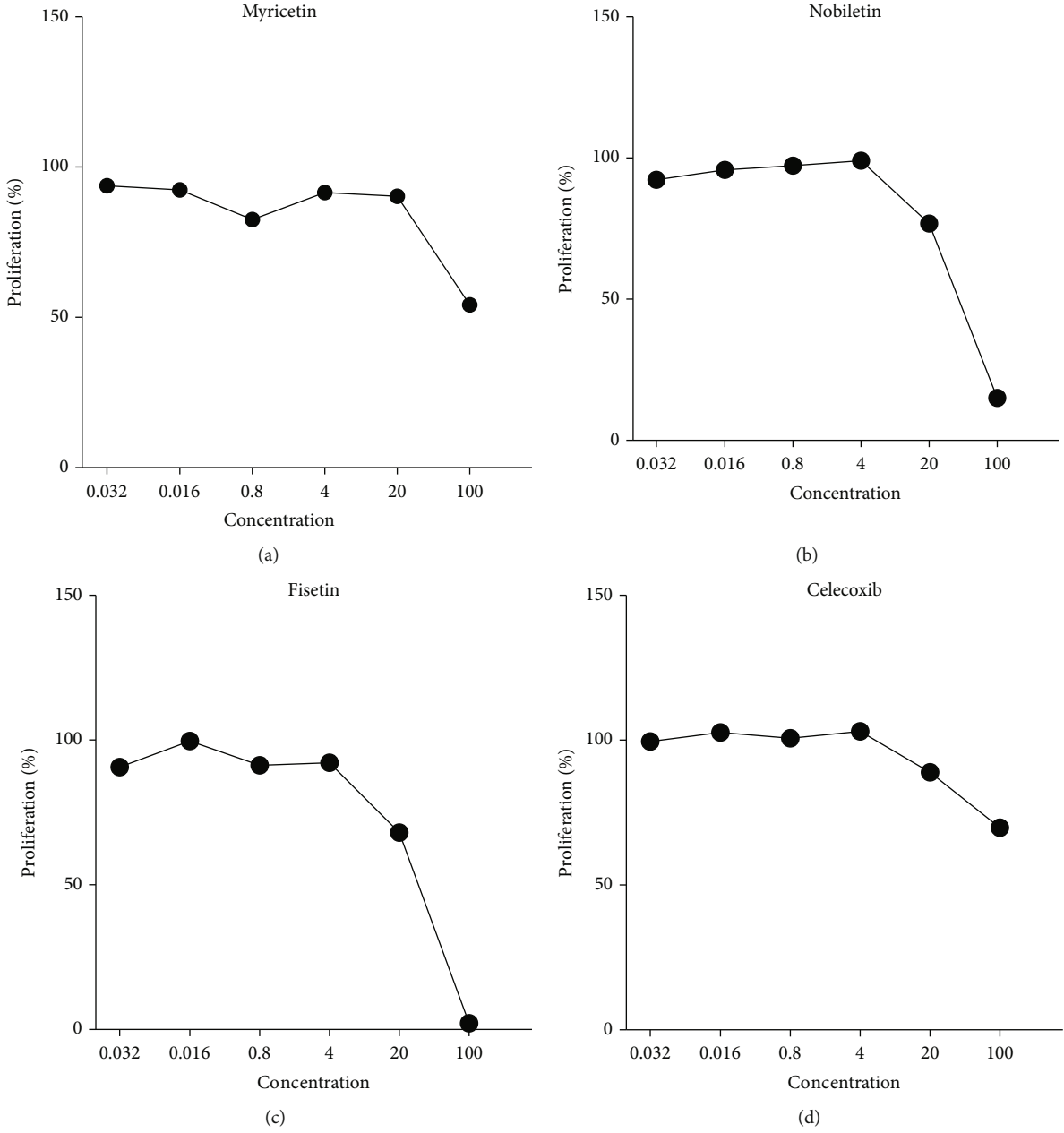
(a)



(b)



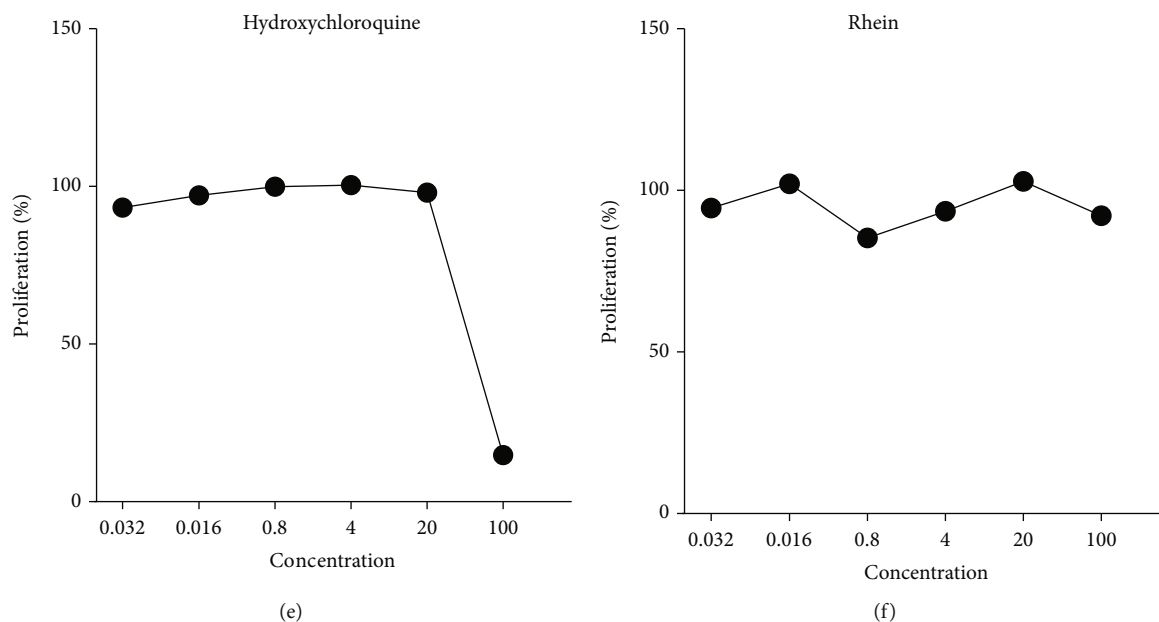(c)



(d)

Figure 6: Continued.

(e)



(f)

FIGURE 6: Effects of myricetin, nobiletin, fisetin, celecoxib, hydroxychloroquine, and rhein on the viability of RAW264.7 cells.

drugs were combined in pairs to obtain potential drug combinations: $C_{41}^1 \times C_{10}^1 = 410$ combinations. Thus, we obtained the final prediction sample set (Supplementary 3; drug combination prediction dataset of integrated Western and TCM drugs).

The results revealed that out of 1000 random sample predictions, 34, 62, and 114 pairs of Western and TCM drugs were predicted to be combination drugs more than 500, 100–500, and 1–100 times, and 200 groups were predicted to be noncombination drugs. In this study, we considered predictions of more than 100 times to be potential anti-RA Western and TCM combination drugs, with higher numbers of positive sample predictions indicating a greater potential for drug combination. Therefore, we selected the 15 groups that were predicted more than 700 times for subsequent cell experiments. The combination drug information is displayed in Table 5 (Supplementary 4; prediction results of all Western and TCM combinations).

*3.6. Experimental Verification Results.* Because RAW264.7 cells were used for experimental verification, we first determined whether a single small molecule had an inhibitory effect on RAW264.7 cells. We then tested the Western and TCM drug combinations. Finally, the CI value was calculated using CompuSyn software to determine whether the small molecule combinations of the two drugs were synergetic.

We then applied the method described in Section 2.5. Based on observations of the proliferation effect of the 15 small molecules on the RAW264.7 cells, myricetin, nobiletin, fisetin, celecoxib, and hydroxychloroquine significantly reduced the survival rate of the RAW264.7 cells. The $IC_{50}$ values of the myricetin, nobiletin, fisetin, celecoxib, and hydroxychloroquine were> 100, 23.93, 31.19, >100, and 79.43 $\mu$mol/L, respectively, as presented in Figure 6.

We further tested combinations of Western and TCM containing these five small molecules (a total of seven pairs) based on the prediction results of the machine learning model and calculated the CI values, as displayed in Table 6. We found that of the seven pairs of Western and TCM combinations, five of the pairs had CI values less than 0.7, indicating that these drugs have good synergy.

## 4. Discussion

Machine learning methods have been widely used in many fields of research, such as text classification [23], sentiment analysis [24], breast cancer diagnosis [25], web page classification [26], and image generation [27]. Many researchers have used computational and modeling methods to identify potential combinatorial drugs [17, 28–30], and the use of machine learning to predict drug combinations has notably increased in recent years [31–34]. However, nearly all of this research has focused of combinations of Western drugs. Combinations of Western and TCM drugs have rarely been reported, especially within the context of RA treatment. TCM is a complex system composed of many small molecules with synergistic effects. In recent years, the research of TCM small molecule is still the focus of modern pharmacology. We mainly focused on the prediction of the combination of small molecule with clear pharmacological effects in TCM and Western medicines that have been approved drugs. Therefore, in this study, we examined the small molecule characteristics of TCM ingredients that differed from those in Western combination drugs. We developed a method for complete modeling and prediction to screen anti-RA combination drugs.

For feature variable construction, we selected drug structures and ADME features. However, the results of the statistical analyses and feature variable screening revealed that

TABLE 6: Effect of seven predicted pairs of Western and TCM combinations on RAW264.7 cell viability.

| Combination | TCM ($\mu$mol/L) | Drug ($\mu$mol/L) | IC50 | RAW264.7 cell inhibition rate (%) | CI |
|---|---|---|---|---|---|
| | | | >100 + 100 | | CI < 1 |
| | 100 | 100 | | 39.31 ± 2.92 | CI = 0.37 |
| | 20 | 100 | | 35.68 ± 1.49 | CI = 0.49 |
| Rhein+celecoxib | 4 | 100 | | 38.60 ± 0.80 | CI = 0.39 |
| | 100 | 20 | | 20.08 ± 1.22 | CI = 0.40 |
| | 100 | 4 | | 14.22 ± 1.59 | CI = 0.17 |
| | 20 | 20 | | 20.17 ± 1.49 | CI = 0.40 |
| | | | 8.83 + 100 | | CI < 1 |
| Nobiletin+celecoxib | 100 | 100 | | 96.20 ± 0.24 | CI = 0.01 |
| | 100 | 20 | | 61.50 ± 1.67 | CI = 0.17 |
| | | | 100 + 4 | | CI < 1 |
| | 100 | 100 | | 64.05 ± 0.94 | CI = 0.06 |
| | 20 | 100 | | 37.13 ± 2.13 | CI = 0.46 |
| Myricetin+celecoxib | 4 | 100 | | 37.96 ± 1.58 | CI = 0.41 |
| | 100 | 20 | | 53.92 ± 1.59 | CI = 0.03 |
| | 100 | 4 | | 49.97 ± 0.93 | CI = 0.02 |
| | | | 10.89 + 100 | | CI < 1 |
| Fisetin+celecoxib | 100 | 100 | | 99.06 ± 0.03 | CI = 0.02 |
| | 4 | 100 | | 57.80 ± 1.89 | CI = 0.88 |
| | | | 100 + 57.15 | | CI < 1 |
| Rhein+hydroxychloroquine | 100 | 20 | | 86.88 ± 0.04 | CI = 0.28 |
| Epicatechin+celecoxib | | | >100 + 100 | | CI > 1 |
| Rosmarinic acid+celecoxib | | | >100 + 100 | | CI > 1 |

drug structure similarity and ADME characteristics had no significant effect on the classification of combination drugs. Notably, sequence similarity, GO functional similarity, and the distance of drug targets within the PPI network of the small molecule drug targets were found to be related to the combination of drugs, with the similarity value of the drug target pathway ($S_{\text{Pathway}}$) being identified as a key feature in combination drugs, both statistically and in feature screening. Therefore, the synergistic mechanism of two or more drugs within a pathway may serve as a key reference for future developments of combined drugs.

In constructing the sample training set, we were unable to collect sufficient noncombined drug data (namely the number of negative samples) due to limited research data. Therefore, in the original data, the number of positive samples was only 1% that of negative samples, which created a serious imbalance in the number of positive and negative samples. If the model had been constructed using the original datasets, the disproportionate number of negative samples would have caused overfitting, and the predictions would have been biased toward the classifications of the larger number of samples. This would have greatly reduced the normalizing ability of the models. However, simply reducing the sample size may have excluded potential drug combinations in the 45603 samples.

To effectively and reasonably overcome this challenge, we proposed a method in which multiple training sets were constructed using random sampling performed 1000 times. For each instance of modeling, the positive samples remained unchanged, and a limited number of negative samples were randomly selected from the total negative samples; 1000 training sets were randomly composed. The machine learning model was trained by and predicted the 1000 training sets. The final potential combination drugs were then determined on the basis of the percentage of each sample predicted to be a combination drug (positive) in the 1000 learning sessions. However, because the number of noncombination drugs is generally much larger than that of combination drugs but large deviations in numbers of positive and negative samples would cause prediction deviation in the machine learning model, in this study, the ratio of positive and negative samples was set to 1 : 5. Through these changes, we were able to prevent imperfect data sampling of combined drugs.

In this study, from the prediction and evaluation results of five different machine learning classification models in

1000 random sample training sets, the accuracy, precision, $F1$ score, AUROC, and AUPR values of RF are the highest among the five machine learning models. AdaBoost performs best on recall, followed by SVM and RF. The prediction scores of $k$-NN and NB on these six machine learning evaluation indexes were lower, suggesting that these two models may not be suitable for combination drug prediction. The prediction performance of SVM and AdaBoost is second only to that of RF. Therefore, through the evaluation and comparison of the predictive ability of different models, we found that the RF classification index scores are generally better than other predictive classifiers. Therefore, the RF model is used as the final combined drug prediction classifier, where the parameter $ntree = 300$. We used the RF classifier to predict 410 groups of Western and TCM drug combinations to treat RA and thus preliminarily predicted 96 pairs of potential Western and TCM drug combinations to treat RA. We used RAW264.7 cells to experimentally verify the effects of a single drug small molecule on the proliferation of RAW264.7 cells. We then verified the effects of combined drugs on the cells. The results revealed that five of the seven selected pairs of combined Western and TCM drugs had high synergistic effects, indicating that our model had high accuracy and practicability. The five optimal combinations were Rhein and Celecoxib, Nobiletin and Celecoxib, Myricetin and Celecoxib, Fisetin and Celecoxib, and Rhein and Hydroxychloroquine.

Notably, celecoxib was found to have a higher synergistic therapeutic effect when combined with several small molecule TCM ingredients, namely, myricetin, rhein, nobiletin, and fisetin. In addition, rhein did not demonstrate an inhibitory effect on RAW264.7 cell proliferation when applied alone. However, when combined with celecoxib (CI < 0.7), the combined inhibition rate of 100 $\mu$mol/L rhein and 20 $\mu$mol/L celecoxib was 35.68%, which was higher than the rate of each of the two alone. Rhien had high synergy with hydroxychloroquine.

Notably, triptolide combined with Western drugs, such as methotrexate, has previously been reported to produce favorable results in past experiments [35]. However, the number of predicted combinations containing triptolide in our model was not high; it predicted triptolide combined with diclofenac (75 times), naproxen (75 times), and celecoxib (24 times) but did not predict triptolide combined with methotrexate (0 times). This may be due to triptolide's reported clinical toxicity [36]. Although triptolide has been clinically reported to improve treatment outcomes, it may not be the best option for drug combinations and may even lead to more systemic side effects [37]. Our prediction results may support this hypothesis. The screening of potential combination of TCM and Western medicine based on the machine learning combined with experimental validation methods can improve efficiency and reduce research costs.

In conclusion, we constructed combination drug characteristics based on existing combination drug data and characteristics from TCM drug research. Key characteristic indexes were screened, and a random forest classification model was used to predict potential combinations of Western and TCM drugs for anti-RA treatment. Through experimental verification, we preliminarily demonstrated that this method of combination therapy for anti-RA treatment can be applied to clinical practice. In the future, we will conduct vivo experiments and clinical trials to validate the findings of this study. In addition, our machine learning model for integrated Western and TCM drug combination can be applied for predicting not only anti-RA drug combinations but also TCM combinations for other diseases.

## Abbreviations

| | |
|---|---|
| ADME: | Adsorption, distribution, metabolism, and excretion |
| AUPRC: | Area under the precision–recall curve |
| CAS: | Chemical Abstracts Service |
| CI: | Combination index |
| DMARDs: | Disease modifying antirheumatic drugs |
| FDA: | Food and Drug Administration |
| FN: | False negative |
| FP: | False positive |
| FPR: | False positive rate |
| GO: | Gene ontology |
| KEGG: | Kyoto Encyclopedia of Genes and Genomes |
| $k$-NN: | $k$-nearest neighbors |
| NB: | Naive Bayes |
| NSAIDs: | Nonsteroidal anti-inflammatory drugs |
| PPI: | Protein–protein interaction |
| RA: | Rheumatoid arthritis |
| RF: | Random forest |
| ROC: | Receiver operating characteristic |
| SMILES: | Canonical simplified molecular-input line-entry system |
| SVM: | Support vector machine |
| TCM: | Traditional Chinese medicine |
| TCMSP: | Traditional Chinese Medicine Systems Pharmacology |
| TN: | True negative |
| TP: | True positive |
| TPR: | True positive rate. |

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors have no conflicts of interest to declare regarding the publication of this article.

## Acknowledgments

## Supplementary Materials

*Supplementary 1.* Pretreated FDA-certified combination drug dataset.

*Supplementary 2.* Positive and negative sample training dataset for all drug combinations.

*Supplementary 3.* Drug combination prediction dataset of integrated Western and TCM drugs.

*Supplementary 4.* Prediction results of all Western and TCM combinations.

## References

[1] T. C. Chou, "Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies," *Pharmacological Reviews*, vol. 58, no. 3, pp. 621–681, 2006.

[2] S. U. Bokhari, U. M. Gopal, and W. C. Duckworth, "Beneficial effects of a glyburide/metformin combination preparation in type 2 diabetes mellitus," *The American Journal of the Medical Sciences*, vol. 325, no. 2, pp. 66–69, 2003.

[3] K. Augustin, S. Williams, M. Cunningham et al., "Perampanel and decanoic acid show synergistic action against AMPA receptors and seizures," *Epilepsia*, vol. 59, no. 11, pp. e172–e178, 2018.

[4] Z. Y. Zhang, M. Chen, J. Chen, M. V. Padval, and V. V. Kansra, "Biotransformation and in vitro assessment of metabolism-associated drug-drug interaction for CRx-102, a novel combination drug candidate," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 50, no. 2, pp. 200–209, 2009.

[5] K. Chaudhari, S. Rizvi, and B. A. Syed, "Rheumatoid arthritis: current and future trends," *Nature Reviews. Drug Discovery*, vol. 15, no. 5, pp. 305-306, 2016.

[6] D. A. Walsh and D. F. McWilliams, "Mechanisms, impact and management of pain in rheumatoid arthritis," *Nature Reviews Rheumatology*, vol. 10, no. 10, pp. 581–592, 2014.

[7] I. M. Chung, S. Ketharnathan, M. Thiruvengadam, and G. Rajakumar, "Rheumatoid arthritis: the stride from research to clinical practice," *International Journal of Molecular Sciences*, vol. 17, no. 6, p. 900, 2016.

[8] L. Gossec and M. Dougados, "Combination therapy in early rheumatoid arthritis," *Clinical and Experimental Rheumatology*, vol. 21, no. 5, Supplement 31, pp. S174–S178, 2003.

[9] G. S. Hazlewood, C. Barnabe, G. Tomlinson et al., "Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis: a network meta-analysis," *Cochrane Database of Systematic Reviews*, vol. 2016, no. 8, article Cd010227, 2016.

[10] Z. Xia, J. Lyu, N. Hou, L. Song, X. Li, and H. Liu, "Iguratimod in combination with methotrexate in active rheumatoid arthritis," *Zeitschrift für Rheumatologie*, vol. 75, no. 8, pp. 828–833, 2016.

[11] E. H. Ibfelt, R. K. Jacobsen, T. I. Kopp et al., "Methotrexate and risk of interstitial lung disease and respiratory failure in rheumatoid arthritis: a nationwide population-based study," *Rheumatology*, vol. 60, no. 1, pp. 346–352, 2021.

[12] K. Albrecht and U. Müller-Ladner, "Side effects and management of side effects of methotrexate in rheumatoid arthritis," *Clinical and Experimental Rheumatology*, vol. 28, no. 5, Supplement 61, pp. S95–101, 2010.

[13] W. Wang, X. Wang, X. Tang, Q. Jiang, and Y. Fan, "Classifying rheumatoid arthritis by traditional Chinese medicine Zheng: a multi-center cross-sectional study," *Journal of Traditional Chinese Medicine*, vol. 39, no. 3, pp. 425–432, 2019.

[14] P. Zhang, J. Li, Y. Han, X. W. Yu, and L. Qin, "Traditional Chinese medicine in the treatment of rheumatoid arthritis: a general review," *Rheumatology International*, vol. 30, no. 6, pp. 713–718, 2010.

[15] H. Liu, W. Zhang, B. Zou, J. Wang, Y. Deng, and L. Deng, "DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy," *Nucleic Acids Research*, vol. 48, no. D1, pp. D871–D881, 2020.

[16] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Scientific Reports*, vol. 7, no. 1, article 42717, 2017.

[17] F. Cheng, I. A. Kovács, and A. L. Barabási, "Network-based prediction of drug combinations," *Nature Communications*, vol. 10, no. 1, article 1197, 2019.

[18] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[19] J. Kohonen, S. Talikota, J. Corander, P. Auvinen, and E. Arjas, "A naive Bayes classifier for protein function prediction," *In Silico Biology*, vol. 9, no. 1,2, pp. 23–34, 2009.

[20] T. Zhang, "An introduction to support vector machines and other kernel-based learning methods," *AI Magazine*, vol. 30, p. 103, 2001.

[21] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, pp. 217–222, 2007.

[22] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, article 1612, 1999.

[23] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.

[24] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency and Computation Practice and Experience*, vol. 33, no. 23, article e5909, 2020.

[25] A. Onan, "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer," *Expert Systems with Applications*, vol. 42, no. 20, pp. 6844–6852, 2015.

[26] A. Onan, "Classifier and feature set ensembles for web page classification," *Journal of Information Science*, vol. 42, no. 2, pp. 150–165, 2016.

[27] M. A. Abazari, M. Soltani, F. Moradi Kashkooli, and K. Raahemifar, "Synthetic 18F-FDG PET image generation using a combination of biomathematical modeling and machine learning," *Cancers*, vol. 14, no. 11, p. 2786, 2022.

[28] X. M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort, and P. Bork, "Prediction of drug combinations by integrating

molecular and pharmacological data," *PLoS Computational Biology*, vol. 7, no. 12, article e1002323, 2011.

[29] Y. Y. Wang, K. J. Xu, J. Song, and X. M. Zhao, "Exploring drug combinations in genetic interaction network," *BMC Bioinformatics*, vol. 13, Supplement 7, p. S7, 2012.

[30] H. Huang, P. Zhang, X. A. Qu, P. Sanseau, and L. Yang, "Systematic prediction of drug combinations based on clinical side-effects," *Scientific Reports*, vol. 4, no. 1, article 7160, 2014.

[31] Y. Sun, Y. Xiong, Q. Xu, and D. Wei, "A hadoop-based method to predict potential effective drug combination," *BioMed Research International*, vol. 2014, Article ID 196858, 5 pages, 2014.

[32] X. Li, G. Qin, Q. Yang, L. Chen, and L. Xie, "Biomolecular network-based synergistic drug combination discovery," *BioMed Research International*, vol. 2016, Article ID 8518945, 11 pages, 2016.

[33] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: predicting synergistic drug combinations based on semi-supervised learning," *PLoS Computational Biology*, vol. 12, no. 7, article e1004975, 2016.

[34] H. L. Rabben, G. T. Andersen, A. Ianevski et al., "Computational drug repositioning and experimental validation of ivermectin in treatment of gastric cancer," *Frontiers in Pharmacology*, vol. 12, article 625991, 2021.

[35] W. Li, K. Gong, Y. Ding et al., "Effects of triptolide and methotrexate nanosuspensions on left ventricular remodeling in autoimmune myocarditis rats," *International Journal of Nanomedicine*, vol. 14, pp. 851–863, 2019.

[36] C. Xi, S. Peng, Z. Wu, Q. Zhou, and J. Zhou, "Toxicity of triptolide and the molecular mechanisms involved," *Biomedicine & Pharmacotherapy*, vol. 90, pp. 531–541, 2017.

[37] A. E. Grzegorzewska, D. Frankiewicz, D. Bręborowicz, I. Matławska, and W. Bylka, "Disseminated cutaneous Kaposi sarcoma in a patient receiving triptolide/tripdiolide for rheumatoid arthritis," *Medical Science Monitor*, vol. 18, no. 8, article Cs67, 2012.