Integration of various protein similarities using random forest technique to infer augmented drug-protein matrix for enhancing drug-disease association prediction

Satanat Kitsiranuwat1,2, Apichat Suratanee3,4 and Kitiporn Plaimas2,5
1Program in Bioinformatics and Computational Biology, Graduate School, Chulalongkorn University, Bangkok, Thailand
2Advanced Virtual and Intelligent Computing (AVIC) center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
3Department of Mathematics, Faculty of Applied Science, King Mongkut’s University of Technology North Bangkok, Bangkok, Thailand
4Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut’s University of Technology North Bangkok, Bangkok, Thailand
5Omics Sciences and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

Abstract
Identifying new therapeutic indications for existing drugs is a major challenge in drug repositioning. Most computational drug repositioning methods focus on known targets. Analyzing multiple aspects of various protein associations provides an opportunity to discover underlying drug-associated proteins that can be used to improve the performance of the drug repositioning approaches. In this study, machine learning models were developed based on the similarities of diversified biological features, including protein interaction, topological network, sequence alignment, and biological function to predict protein pairs associating with the same drugs. The crucial set of features was identified, and the high performances of protein pair predictions were achieved with an area under the curve (AUC) value of more than 93%. Based on drug chemical structures,

Corresponding author:
Kitiporn Plaimas, Advanced Virtual and Intelligent Computing (AVIC) center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.
Email: kitiporn.p@chula.ac.th

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
the drug similarity levels of the promising protein pairs were used to quantify the inferred drug-associated proteins. Furthermore, these proteins were employed to establish an augmented drug-protein matrix to enhance the efficiency of three existing drug repositioning techniques: a similarity constrained matrix factorization for the drug-disease associations (SCMFDD), an ensemble meta-paths and singular value decomposition (EMP-SVD) model, and a topology similarity and singular value decomposition (TS-SVD) technique. The results showed that the augmented matrix helped to improve the performance up to 4% more in comparison to the original matrix for SCMFDD and EMP-SVD, and about 1% more for TS-SVD. In summary, inferring new protein pairs related to the same drugs increase the opportunity to reveal missing drug-associated proteins that are important for drug development via the drug repositioning technique.

**Keywords**
Protein-protein interaction network, drug repositioning, drug repurposing, machine learning, drug-protein association

**Introduction**

The similarities among proteins have been applied to many bioinformatics fields, including drug repositioning. The indication of a drug can be repurposed to treat another disease by targeting other similar target proteins. However, there are no gold standard features to identify the similarities among proteins to repurpose the use of drugs. Therefore, investigating various features to describe proteins sharing common drugs can be beneficial in a drug repositioning technique.

Drug repositioning, also known as drug repurposing, is a technique for using existing drugs for new indications. In principle, the procedure of drug discovery involves i) screening and searching for the compounds associated with the disease in the laboratory, ii) confirmation of safety for indication uses, iii) clinical research phase I, iv) clinical research phase II, v) clinical research phase III to finally confirm the drug usage on people, vi) approval of the drug by Food and Drug Administration (FDA) reviews, and vii) FDA post-market safety to ensure public availability of the drug. As the chemical structures of existing drugs are already known and these drugs are safe for humans, finding new indications of existing drugs can reduce the cost, resources, and time required to find a new drug to treat disease. For example, Pfizer discovered Sildenafil for curing coronary artery disease. This drug has also been known as Viagra since 1989 whose usage is repurposed to treat erectile dysfunction by increasing blood flow to the penis.

In drug repositioning, the identification of drug-disease associations is the first step to screen common existing drugs to cure various diseases. However, the in vitro experiments for identifying drug-disease associations are time intensive and costly. Therefore, the computational approaches of drug repositioning to predict novel drug-disease associations have become an important task. One of the most important steps to infer new associations is the way to define common middle information that might be related or linked to both drugs and diseases. This common link can be a protein that is related to a certain drug or disease. Drug-protein associations play an important role in the identification of drug-disease associations. Drug-protein associations represent the binding between drugs and proteins to utilize poly-pharmacology concepts such as chemical substructures, pharmacophore functional sites, and pathways. The identification
of possible drug-protein target interactions is a vital procedure in accelerating drug development and drug design; particularly, in drug repositioning, since it reduces the number of chemical compounds that potentially bind to potential targets.\textsuperscript{11,12} The associations between drugs and protein targets play an important role to disclose the functions and chemical structures of pharmaceutically protein targets such as enzymes, ion channels, G-protein-coupled receptors (GPCRs), and nuclear receptors including structure similarity and sequence similarity between drugs and their targets.\textsuperscript{13} The computational technique for drug repositioning is effective in predicting and guiding that an existing drug can be used to cure other diseases as well as to treat drug-resistant cases. Zhang et al. proposed the computational model of a similarity constrained matrix factorization for the drug-disease associations (SCMFDD) based on known drug-disease associations from a curated database, drug features, and disease semantic descriptors.\textsuperscript{9} Protein target was utilized to account for computing drug-drug similarities features including substructures, enzymes, pathways, and drug-drug interactions while disease semantic information was calculated from MeSH information.

Several computational models have been developed for drug repositioning based on the similarities among target proteins. These techniques are established based on a similarity scheme, called guilt-by-association,\textsuperscript{14} which can be expressed in many ways to define the similarities between two drugs and among the target proteins, including disease proteins. The similarity scheme is applied to predict drug-target interactions to support the drug repositioning approach.\textsuperscript{15–17} Gottlieb et al. proposed the PREDICT method to predict drug-disease indications based on the similarity of the drugs that are used to cure similar diseases.\textsuperscript{1} This method combines the two schemes of the drug similarities and disease similarities. In the case of a drug similarities scheme, the authors investigated the information regarding protein-protein interaction (PPI) data, gene ontology, sequence alignment, and phenotypes from target-related drugs. Afterward, they ensembled the two similarity schemes using geometric means to represent the maximum similarity score for each candidate drug-disease pair. They obtained an area under the curve (AUC) of 0.92 for predicting drug indications. Zhang et al. introduced the Similarity-based LArge-margin learning of Multiple Sources (SLAMS) method based on the multiple data sources of a drug’s chemical structures, protein targets and side-effect profiles to retrieve novel drugs for diseases.\textsuperscript{2} The SLAMS method integrates various similarity-based features of drugs and protein targets, which can play an important role in the drug design, and are related to therapeutic use for helping in the drug repositioning model. With the same integrated multiple data sources, the SLAMS method achieved an AUC of 0.89, while the PREDICT method achieved an AUC of 0.87. Khalid et al. proposed the similarity scheme that predicts approved and novel drug targets with new disease associations (SPANTD) method,\textsuperscript{7} which reveals various interesting features that can be combined into a scoring matrix. These features include similarity among proteins, similarity among proteins’ module pathways in the biological function, the pairwise binding site’s structural similarity among proteins, and disease-disease similarity. The authors combined all similarity features with a scoring matrix for drug disease associations. Later, they applied a genetic algorithm to compute the scoring matrix and then predicted the drug-disease association. The SPANTD method achieved an AUC of 0.97 to predict candidate drug-disease associations.
Network-based methods have been successfully used for predicting several tasks, including disease-disease association predictions,\textsuperscript{18} disease protein association predictions,\textsuperscript{19–25} and drug-disease association predictions.\textsuperscript{26,27} Several computational drug repositioning approaches focus on a heterogeneous network of different types of nodes such as drugs, proteins, and diseases. Wu et al. proposed the ensemble meta-paths and singular value decomposition (EMP-SVD) model, which generates five meta-paths, and constructed the latent features of drugs and diseases using the singular value decomposition (SVD) technique.\textsuperscript{28} They investigated the reliable negative, which is the set of drugs that cannot treat diseases. Then, for each meta-path, they employed a random forest algorithm to construct a classifier corresponding to each path. All five classifiers were ensembled to predict candidate drugs for new indications. Moreover, the EMP-SVD method was improved to a new version called topology similarity and singular value decomposition (TS-SVD).\textsuperscript{29} This method integrates the common neighbors count matrices of drugs and diseases constructed based on a heterogeneous network to achieve topological similarity matrices of drugs and diseases. After that, the dimension of the topological similarity matrix of drug and disease was reduced using SVD to represent drug-disease pairs. Then, the authors employed a random forest classifier to predict potential drug-disease associations based on the reliable negative, which is defined by the $k$-step neighbors among drugs and diseases. Both the EMP-SVD and TS-SVD models were generated based on proteins, which were presented as the middle nodes to link drugs and diseases in a heterogeneous network. Another technique that utilized the function of proteins or gene ontology (GO) profiles as the middle nodes to link drugs and diseases in the tripartite network, called meta-path-based gene ontology profiles for predicting drug-disease associations (MGP-DDA), was proposed.\textsuperscript{26} The MGP-DDA model integrates a meta-path based on GO terms to construct a drug repositioning model.

Recently, the protein-protein similarity vectors (PPSVs) technique has been proposed to develop drug repositioning based on the multi-data aspects of protein similarity, such as network topology, proteomic data, functional analysis, and druggable property to determine the associations between proteins and their approved drug.\textsuperscript{27} The PPSVs exploited the manner of separating the drug-disease matrix for individual diseases to emphasize the potential drugs that can treat a specific disease. Then, the random forest classifier was applied to predict candidate drug-disease associations. The PPSVs achieved an AUC of 98.9%. As several previous studies have used the scheme of similarities among target proteins to predict drug-disease pairs, the similarity of protein interactions might also be crucial to explain the common drugs among the target proteins.

In this study, the relationship among drug-associated proteins based on several biological aspects is applied to drug repositioning approaches. The similarities of protein pairs that are related to the same drugs can play a potential role in the drug repositioning technique. The prediction of the associated proteins sharing the same drugs is performed by a machine learning technique, and the prediction results are used to create an augmented drug-protein matrix that enhances the efficiency of existing drug repositioning models. Section 2 characterizes the proposed method for investigating protein pairs that share the same drugs and the process of achieving an augmented drug-protein matrix. Section 3 demonstrates the performance of predicting protein pairs associating
the same drugs. This section also compares the performances of the original drug-protein and the augmented drug-protein matrices when subjected to drug repositioning approaches. Section 4 discusses the results. Section 5 concludes the study and presents future work.

**Materials and methods**

The overview of this study is summarized in five steps, as shown in Figure 1. First, data on drugs, diseases, gene ontologies, protein interactions, and drug-disease interactions are collected based on several published databases. Second, feature analysis is performed by feature selection in different biological meanings categorized into four groups: (1) protein interactions, (2) topological network, (3) sequence alignment, and (4) similarity of functions among proteins. Third, the prediction process is established from ensemble random forest classifiers. Fourth, an augmented drug-protein matrix is inferred based on the target protein prediction. Finally, the drug-disease associations were predicted using the existing drug repositioning models with the augmented drug-protein matrices.

**Dataset**

The human (Homo sapiens) proteins were fetched from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING v.11) database. The ensemble protein IDs in
STRING were converted to protein symbols by using the Universal Protein Resource (Uniprot) database. This mapping resulted in 3,591,273 interactions of human proteins. The DrugBank database was employed to identify the approved drugs and their target proteins as well as their genes. Only the published evidence of the approved drugs and their target proteins was considered from the DrugBank database.

The human protein pairs having common approved drugs were assigned as positive labels. Otherwise, the protein pairs were assigned negative labels. In total, we obtained 27,683 and 3,563,590 human protein pairs for the positive and negative labels, respectively. The protein sequences were retrieved from uniprot.org by protein symbols using the getUniProt in the R language. The level modules for understanding functions in biological systems were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. GO annotation was achieved from the Gene Ontology Annotation (GOA) database. The computational drug similarity was retrieved from DD.chem.data, represented in the bionetdata package v1.0.1 of R language. The DD.chem.data provides the scoring matrix, which represents the chemical structure similarity among the approved drugs from the DrugBank database. The scoring matrix contains the real value, where zero represents a completely different chemical structure between drugs, and one means an identical chemical structure between drugs.

In drug repositioning approaches, information on drug-disease interactions was extracted from the Comparative Toxicogenomics Database (CTD). Moreover, the set of genes or proteins that interact with diseases originated from a database of gene-disease associations (DisGeNET).

**Features analysis**

We observed similarity among proteins in several biological aspects categorized into four different groups: PPI, topological network, sequence alignment, and biological functions. We obtained a total of 13 features, as described below:

**PPI data.** Seven features of protein interactions were retrieved from the STRING v.11 database.

- Conserved neighborhood: This is the inter-gene nucleotide counting that occurs repeatedly in a close neighborhood of genomes.
- Fusion: The score is obtained from individual gene-fused events in other species.
- Co-occurrence: This is derived from the presence or absence of similar patterns of genes in the phylogenetic profile.
- Co-expression: The score is computed from similar patterns of mRNA expression levels in the same or other species.
- Experiments: This is the list of significant protein interactions obtained from affinity chromatography.
- Databases: The score is derived from various curated databases.
- Text mining: This is computed from the co-occurrence of genes or protein names in the abstract of scientific literature.
Topological network data. To extract topological network data, a PPI network model was constructed by retrieving protein interactions from the STRING database. Only interactions with high confidence scores (more than 800) were selected. Then, a weighted adjacency matrix of this was used to calculate two topological features which were common neighbors and closer proteins as follows.

- **Common neighbors**: The common neighbor score, $Nei(P_j, P_k)$, represents the similarity of common neighbors between the two proteins, $j$ and $k$, in the network. It can be calculated using the cosine similarity as follows:

$$Nei(P_j, P_k) = \frac{N(P_j) \cdot N(P_k)}{||N(P_j)|| ||N(P_k)||} = \frac{|N(P_j) \cap N(P_k)|}{\sqrt{|N(P_j)| \cdot |N(P_k)|}} \quad (1)$$

where $N(P_j)$ and $N(P_k)$ are the neighborhood vectors in proteins $j$ and $k$, respectively. This score lies between zero and one. Zero means no common neighbors between the two proteins, while one means that all proteins are the protein neighbors.

- **Closer proteins**: The closer protein score, $Closer(P_j, P_k)$, was calculated from the inverse of the length of the shortest path between two proteins in the network. This can be computed as follows:

$$Closer(P_j, P_k) = \frac{1}{D(P_j, P_k)} \quad (2)$$

where $D(P_j, P_k)$ is the step of the shortest path between proteins $j$ and $k$. The score lies between zero and one. The score for self-proteins is one, while that for two disjoint proteins is zero.

Sequence alignment data. The similarity of proteins’ sequences regions was identified. The similarity between the two sequences can conserve the structure, function, or evolution of these proteins. Two approaches, local and global alignments, were used to identify similarities.

- **Local alignment**: The local alignment score represents the alignment of the most similar regions between the two protein sequences. The local alignment method employs BLOSUM62 for the substitution matrix. The gap opening is assigned a value of 10, while the gap extension is assigned 0.5.

- **Global alignment**: The global alignment score represents the alignment of similarities in the whole sequences of any two proteins. The global alignment parameter is the same substitution matrix as that used in the local alignment approach.

Biological function data. The biological functions between proteins in terms of the similarity of modules, including GO domains, were integrated. The common modules and GO are detailed as below.
Common modules: The common module, \( PW(P_j, P_k) \), is computed from the number of common modules between two proteins. The protein modules are obtained from the KEGG database, which is the collection of the molecular interaction, reaction, and relation.\(^{34-36} \) The score can be computed as follows:

\[
PW(P_j, P_k) = |PW(P_j) \cap PW(P_k)|
\]  

where \( PW(P_j) \) and \( PW(P_k) \) are the module sets in proteins \( j \) and \( k \), respectively.

GO: The gene ontology, \( GO(P_j, P_k) \), represents the similarity ontology domains of the cellular component, molecular function, and biological process between two proteins. It was retrieved from the GOA database.\(^{37} \) The score can be calculated based on cosine similarity as follows:

\[
GO(P_j, P_k) = \frac{|GO(P_j) \cap GO(P_k)|}{\sqrt{|GO(P_j)| \cdot |GO(P_k)|}}
\]  

where \( GO(P_j) \) and \( GO(P_k) \) are the sets of GO profiles in proteins \( j \) and \( k \), respectively.

This study attempted to collect several related biological features for protein pairs. All biological features were rescaled to standard values between zero and one. However, some of these features might be irrelevant to classifying protein pairs that share common drugs. To verify the features, a feature selection technique was employed to retrieve the crucial features for prediction. These are detailed as follows:

Forward selection: This method starts with no feature and then continues adding the most relevant feature which improves the performance value of the AUC score. The method repeatedly adds the feature until the added feature can no longer improve the performance value.

Predicting protein pairs associated with the same drugs

This study employed a random forest classifier to predict protein pairs sharing the same drugs. A grid search technique was utilized to identify the best hyperparameters in the forest. The parameters were set in increments of 50 from 50 to 300 for the number of trees and were set as the square root of the number of features to examine the best split. The vector of label classes is assigned as a binary vector, in which one represents a positive label (protein pairs that have common approved drugs) and zero represents a negative label.

The framework for determining crucial features is illustrated in Figure 2. First, the data of protein pairs are randomly split, with 20% for a test set and the remaining for a training set. To split the protein pairs data into a training set and a test set, the proportion of positive and negative labels remains with the same proportion in the protein pair data. Second, the negative labels are randomly selected to obtain the balance data between the positive and negative labels for both a training set and a test set. Third, the protein pairs in the training set are randomly split, with 80% for generating a random forest classifier based on a forward selection and 20% for validating the performance of forward selection. Fourth, 20% of the test set is applied to evaluate the performance of the optimal
feature model in one experiment. Later, all four steps above are repeated in 10 experiments to prevent bias from randomly splitting the data. After that, 10 optimal feature sets are obtained from 10 experiments. Then, all 10 optimal feature sets are voted to achieve crucial features. Next, the crucial features are applied to a random forest model to predict protein pairs sharing common drugs. The protein pair data are randomly split, with 20% for a test set and the remaining for a training set. A random forest model with crucial features obtained from the training set was generated. The performance of the model was evaluated on the test set. These training and testing processes were performed for five iterations, and the average performance was evaluated. Finally, the protein pairs with an average score exceeding 0.5 were inferred to share common drugs.

**Generating an augmented drug-protein matrix**

This study focuses on drug-protein associations to reinforce the performance of the existing drug repositioning approaches such as EMP-SVD and TS-SVD. The protein pairs predicted to share common drugs were employed to augment the drug-protein matrix used in the approaches. The EMP-SVD and TS-SVD methods were used based on a heterogeneous network with three node types: drugs, proteins, and diseases. The edges represent interactions between nodes, including drug-protein, drug-disease, and protein-disease interactions. The data on approved drug-protein interactions were insufficient. Therefore, the prediction of protein pairs sharing common drugs was applied to generate an augmented drug-protein matrix (ADPM). If protein A and protein B were predicted to share common drugs, then protein A might be related to drugs with target protein B. In contrast, protein B might be related to drugs with target protein A. Then, ADPM can be
calculated as follows:

\[
ADPM(d, p) = \begin{cases} 
1; & \text{if } Pred(p_i, p_j) \geq 0.5 \text{ or } d = Dr(p), \\
0; & \text{otherwise}
\end{cases}
\]  

(5)

where \(d \in \{Dr(p_i), Dr(p_j)\}\), \(p \in \{p_i, p_j\}\), \(Dr(p_i)\) and \(Dr(p_j)\) are the sets of drugs for proteins \(i\) and \(j\), respectively, and \(Pred(p_i, p_j)\) is the prediction score of proteins \(p_i\) and \(p_j\) from the previous section.

**Evaluating performance**

In this study, evaluation metrics of classification performance include AUC, accuracy, precision, sensitivity, F1, and area under precision-recall curve (AUPR) scores. The performance scores can be explained as follows.

The AUC value is the area under the curve of the receiver operating characteristic (ROC) which is plotted between sensitivity (\(SEN\)) and false positive rate (\(FPR\)) at different thresholds. \(SEN\) and \(FPR\) are calculated as follows:

\[
SEN = \frac{TP}{TP + FN}  
\]  

(6)

\[
FPR = \frac{FP}{FP + TN}  
\]  

(7)

where \(TP\) is the number of correctly predicted protein pairs found in the positive set, \(TN\) is the number of correctly predicted pairs found in the negative set, \(FN\) is the number of positive protein pairs incorrectly predicted as negative, and \(FP\) is the number of negative protein pairs incorrectly predicted as positive.

The accuracy (\(ACC\)) is the overall correct prediction, which can be computed as follows:

\[
ACC = \frac{TP + TN}{TP + TN + FP + FN}  
\]  

(8)

The F1 score is the harmonic mean of precision (\(PRE\)) and sensitivity (\(SEN\)), which are expressed in Equations (9) and (6), respectively. The F1 score can be computed as follows:

\[
PRE = \frac{TP}{TP + FP}  
\]  

(9)

\[
F1 = \frac{2 \times PRE \times SEN}{(PRE + SEN)}  
\]  

(10)

In addition, AUPR, which is the area under the curve of plotting between precision on the y-axis and sensitivity on the x-axis for several thresholds of the prediction score, is computed to visualize the model performance. All performance scores range between zero and one. A performance score close to one indicates high prediction accuracy.
Results
This section demonstrates the performance outcomes of predicting protein pairs that might have common drugs based on crucial features. Moreover, the analysis of drugs’ similarities between protein pairs is considered. In addition, we apply the protein pairs that might have common drugs to generate an augmented drug-protein matrix to enhance the performance of the EMP-SVD, an existing drug repositioning approach.

Identification of optimal features to classify protein pairs sharing common drugs
To obtain significant features for identifying protein pairs sharing common drugs, we employed a forward selection technique to determine the crucial features. This technique starts adding the most relevant features for prediction and then adds the other relevant feature. These processes are performed continuously in each step until the algorithm cannot find any features that can improve performance. In our experiments, we used AUC as our performance measure. We performed overall 10 experiments to obtain the set of optimal features. All procedures are shown in Figure 3. Local alignment was used as an initial feature in all 10 experiments; the results showed an AUC greater than 0.93. To perform the majority votes, we observed whether a feature was visible in all sets of optimal features. Interestingly, the fusion feature was found in only three sets of optimal features. The database and co-occurrence features were found in seven and eight sets of optimal features, respectively. The remaining features were found in all 10 sets of optimal features, respectively. Therefore, we discarded only the fusion feature. The results are shown in Figure 4. Finally, we obtained 12 optimal features: (1) local alignment, (2) GO, (3) experiments, (4) common modules, (5) text mining,

| Experiment | ACC  | AUC  | AUPR | F1   | PRE  | SEN  |
|------------|------|------|------|------|------|------|
| Experiment 1 | 0.8576 | 0.9380 | 0.9323 | 0.8665 | 0.8154 | 0.9245 |
| Experiment 2 | 0.8596 | 0.9401 | 0.9351 | 0.8672 | 0.8225 | 0.9171 |
| Experiment 3 | 0.8590 | 0.9402 | 0.9332 | 0.8696 | 0.8089 | 0.9042 |
| Experiment 4 | 0.8569 | 0.9362 | 0.9280 | 0.8685 | 0.8035 | 0.9449 |
| Experiment 5 | 0.8510 | 0.9390 | 0.9327 | 0.8625 | 0.8006 | 0.9348 |
| Experiment 6 | 0.8570 | 0.9368 | 0.9312 | 0.8675 | 0.8082 | 0.9362 |
| Experiment 7 | 0.8559 | 0.9379 | 0.9324 | 0.8667 | 0.8062 | 0.9370 |
| Experiment 8 | 0.8655 | 0.9446 | 0.9414 | 0.8717 | 0.8335 | 0.9135 |
| Experiment 9 | 0.8589 | 0.9415 | 0.9379 | 0.8676 | 0.8175 | 0.9241 |
| Experiment 10 | 0.8538 | 0.9369 | 0.9312 | 0.8649 | 0.8037 | 0.9362 |

Figure 3. Performance of 10 experiments using the forward selection technique.
(6) global alignment, (7) co-expression, (8) common neighbors, (9) conserved neighborhood, (10) closer proteins, (11) databases, and (12) co-occurrence. These crucial features were used by the random forest classifiers to predict protein pairs sharing common drugs. The optimal features in each experiment, as well as their performance values, are shown in Supplementary Table S1.

**Performance of predicting protein pairs sharing common drugs**

The crucial features obtained from the forward selection method were used by the random forest classifiers. Classifications with the random forest model were performed for five iterations with random sets of training and test data. Table 1 shows the classification performances and their average values. The results showed an average accuracy score of 85.7%, representing an overall correct prediction. The average precision and sensitivity of the model approximated as 0.813 and 0.928, respectively. This means that the model correctly predicted positive labels from all positive predictions with 81.3% accuracy and actual positives with 92.8% accuracy. The F1 score, which is the harmonic mean

| Iterations | AUC    | AUPR   | PRE    | SEN    | ACC    | F1     |
|------------|--------|--------|--------|--------|--------|--------|
| 1          | 0.938042 | 0.932256 | 0.815358 | 0.924494 | 0.857569 | 0.866503 |
| 2          | 0.940308 | 0.935017 | 0.816502 | 0.925939 | 0.858923 | 0.867784 |
| 3          | 0.939822 | 0.932861 | 0.802847 | 0.947435 | 0.857388 | 0.869169 |
| 4          | 0.935980 | 0.927651 | 0.817325 | 0.923772 | 0.858652 | 0.867294 |
| 5          | 0.938990 | 0.932777 | 0.811543 | 0.919436 | 0.852962 | 0.862127 |
| Average    | 0.938628 | 0.932112 | 0.812715 | 0.928215 | 0.857099 | 0.866575 |

**Figure 4.** The number of experiments in which the observing feature was found.

**Table 1.** Performance of predicting protein pairs for five iterations.
of precision and sensitivity showed an average value of 0.867. In addition, the area under the ROC curve was 0.939. The ROC curve for the five iterations is shown in Figure 5. There is more than 93% chance of the model classifying positive and negative labels for all five iterations. Moreover, the average AUPR is 0.932. The precision-recall curve for five iterations is illustrated in Figure 6.

**Novel drug associations corresponding to the predicted protein pairs**

The average prediction score of each protein pair was calculated from the prediction scores of the pairs from all five classification models. If the average prediction score was greater than 0.5, then the protein pair was inferred to share common drugs. However, the amount of positive data was dominated by that of negative data (positive labels: 27,683 and negative labels: 3,563,590). Positive data can be unrevealed in the negative data group. Our results indicated 638,830 false positive pairs which is much higher than the number of true positive pairs (27,479). Therefore, we investigated the group of false positive prediction pairs using the similarity levels of the drug’s chemical structures. These levels were derived from the maximum values of the chemical structure similarity of all possible drug pairs for a protein pair using the DD.CHEM database. The highest level of drug similarity was one, which indicates an identical chemical structure between two drugs. A higher level of drug similarity indicates the similarity of their chemical structures. Therefore, our predicted protein pairs with high drug similarity

![Figure 5. ROC curve for five iterations.](image-url)
levels can presumably be proteins that are related to the same drugs. In the false positive group, there are 4718 pairs with drug similarity levels greater than or equal to 0.5. A quarter of these pairs have a level $\geq 0.7$. There are up to 85 protein pairs having the highest level of 1. This is a hint that several protein pairs are yet to be discovered to share common drugs. All protein pairs having drug similarity levels greater than or equal to 0.5 are reported in Supplementary Table S2. Table 2 shows an example of 10 protein pairs with the highest drug similarity level of 1.

**Table 2.** Ten protein pairs with the highest drug similarity levels.

| Target protein 1 | Target protein 2 | Drug 1    | Drug 2    |
|------------------|------------------|-----------|-----------|
| P25021           | P13945           | DB00751   | DB00368   |
| P07550           | P41145           | DB06216   | DB00295   |
| P35372           | P08588           | DB00611   | DB00408   |
| P35372           | P07550           | DB00704   | DB00397   |
| P13945           | P41145           | DB01001   | DB00295   |
| P08588           | P41143           | DB11273   | DB00295   |
| P35372           | P13945           | DB00921   | DB00368   |
| P18825           | P41143           | DB01049   | DB00295   |
| P13945           | P41143           | DB11278   | DB00295   |
| P35372           | P18825           | DB00704   | DB00397   |

![Figure 6. Pr curve for five iterations.](image)

levels can presumably be proteins that are related to the same drugs. In the false positive group, there are 4718 pairs with drug similarity levels greater than or equal to 0.5. A quarter of these pairs have a level $\geq 0.7$. There are up to 85 protein pairs having the highest level of 1. This is a hint that several protein pairs are yet to be discovered to share common drugs. All protein pairs having drug similarity levels greater than or equal to 0.5 are reported in Supplementary Table S2. Table 2 shows an example of 10 protein pairs with the highest drug similarity level of 1.
We further investigated the similarity between the two drugs, as shown in Table 2. We employed the CTD database, which reports inferred associations based on chemical-gene interactions. If gene A has a curated association with chemical C and that with disease B, then chemical C has been reported as an inferred association with disease B. The gene that links the drug and disease of the inferred association is called the inferred gene. The inferred associations, on the other hand, report a relationship between a chemical and a condition; it does not imply that the drug has a potential therapeutic role in that disease. The database indicates an inferred association between epinastine (DB00751) and COVID-19 (MESH: C000657245)\textsuperscript{43} and an inferred association between norepinephrine (DB00368) and COVID-19 (MESH: C000657245).\textsuperscript{43–45} Directed evidence of marker association was found. Asenapine (DB06216) and morphine (DB00295) are related to basal ganglia disease (MESH:D001480).\textsuperscript{46,47}

Additionally, both butorphanol (DB00611) and loxapine (DB00408) have inferred associations with Arrhythmias, cardiac disease (MESH:D001145).\textsuperscript{48–51} The database also provides the inferred associations between naltrexone (DB00704) and amyotrophic lateral sclerosis 1 (MESH:C531617),\textsuperscript{52} as well as those between phenylpropanolamine (DB00397) and amyotrophic lateral sclerosis 1 (MESH:C531617).\textsuperscript{52–70}

The drugs that have been reported as inferred associations with COVID-19 (MESH: C000657245) are albuterol (DB01001), morphine (DB00295), buprenorphine (DB00921), and norepinephrine (DB00368).\textsuperscript{43–45,71} In addition, the chemical morphine (DB00295) is marker-directed evidence for vertigo disease (MESH: D014717),\textsuperscript{72,73} while dihydroergocornine (DB11273) is its therapeutic-directed evidence.\textsuperscript{74} Moreover, naltrexone (DB00704) and phenylpropanolamine (DB00397) have inferred associations with esophageal squamous cell carcinoma (MESH:D000077277).\textsuperscript{75,76} For the drug pair between ergoloid mesylate (DB01049) and morphine (DB00295) and the pair between DL-methylephedrine (DB11278) and morphine (DB00295), we could not find evident support for their associations. The full list of inferred associations between drug and disease with inferred gene symbols and the literature support for the drug-drug pairs listed in Table 2 is shown in Supplementary Table S3.

**Augmented drug-protein matrix**

Drug-protein associations can be drawn from various aspects such as direct targets and functional relations. However, the limitation of the associations between drug and protein targets affects the efficiency of inferring new indications of drug repositioning approaches. In this case, the number of protein targets of drugs in the DrugBank database was investigated and revealed that one drug is mostly associated with one protein, as shown in Figure 7. To achieve the aim of drug repositioning, we attempted to find more proteins associated with the drugs. Therefore, our analysis addressed the associations between the proteins and the approved drugs.

The prediction results of protein pairs sharing the same drugs were employed to further obtain more drug-protein associations. Using the criteria in Equation (5), the number of drug-protein associations in $ADPM$ increased from 16,868 associations in the original matrix to 926,204 associations in the augmented matrix. The number of drugs and the number of their target proteins are shown in Figure 8.
Performance of drug-repositioning approaches with the augmented drug-protein matrix

To demonstrate the enhanced efficiency of the drug-repositioning approaches with the use of the augmented drug-protein matrix, SCMFDD, EMP-SVD, and TS-SVD were implemented using the augmented drug-protein matrix. Table 3 shows the performance of each method with the use of the original drug-protein matrix and the augmented drug-protein matrix. These three methods yielded better performance in all performance measures with the use of the augmented drug-protein matrix. TS-SVD with the augmented matrix provided the best performance.

**Figure 7.** The frequency of drugs and their associated proteins are based on direct target protein information from all approved drugs.

**Figure 8.** Frequency of drugs and their associated proteins based on our prediction results of protein pairs related to the same drugs.

**Performance of drug-repositioning approaches with the augmented drug-protein matrix**

To demonstrate the enhanced efficiency of the drug-repositioning approaches with the use of the augmented drug-protein matrix, SCMFDD, EMP-SVD, and TS-SVD were implemented using the augmented drug-protein matrix. Table 3 shows the performance of each method with the use of the original drug-protein matrix and the augmented drug-protein matrix. These three methods yielded better performance in all performance measures with the use of the augmented drug-protein matrix. TS-SVD with the augmented matrix provided the best performance.
In SCMFDD, the augmented drug-protein matrix was modified by converting the drugs’ name from ‘DrugBank ID’ to ‘Chemical MeSH ID’ for enhancing the possibility of drug-protein associations in the drug-protein matrix to compute drug-drug similarities based on Jaccard similarity in SCMFDD method. The SCMFDD method utilized drug feature similarities and disease semantic similarity in low-rank spaces using the matrix factorization technique. The disease semantic similarity was computed from MeSH information. There are five drug features into account drug feature-based similarities such as substructures, protein targets, pathways, enzymes, and drug-drug interactions. Only protein target features can be improved with the augmented drug-protein association. Since the SCMFDD method individually employed drug feature-based similarities to generate predictive models, the result in Table 3 shows the performance of the SCMFDD using the protein target features to compute the drug feature-based similarities. The dataset of the SCMFDD method contains 269 drugs and 598 diseases with 18,416 known drug-disease associations. For drug feature-based dataset composes 881 types of compound substructures, 623 protein targets, 247 enzymes, 465 pathways, and 2086 interactions among drugs. Consequently, the dataset of drugs and proteins from the augmented drug-protein matrix were mapped into the same dataset of the SCMFDD method. The dimensions of the original drug-protein and augmented drug-protein matrix for the SCMFDD method were 269×529. The original drug-protein matrix of SCMFDD has 1,526 associations between drugs and proteins while the number of drug-protein associations increased to 39,420 associations in the augmented drug-protein matrix of the SCMFDD method. The performance of SCMFDD with the augmented drug-protein matrix in Table 3 was superior to that of SCMFDD with the original drug-protein matrix when using protein targets to compute drug feature-based similarity.

The augmented drug-protein matrix was applied to enhance the performance of the existing drug repositioning models (EMP-SVD and TS-SVD). Both models required information on the drug-protein matrix, disease-protein matrix, and drug-disease matrix. All matrices were constructed using various databases such as DrugBank, CTD, and DisGeNET. Consequently, the dimensions of the drug-protein matrix and augmented drug-protein matrix were 2,120×9,314, and those of the drug-disease and disease-protein matrices were 2,120×1,437 and 1,437×9,314, respectively.

### Table 3. Performance of the drug-repositioning techniques with the use of the augmented drug-protein matrix.

| Methods      | Original drug-protein matrix | Augmented drug-protein matrix |
|--------------|-----------------------------|-------------------------------|
|              | AUPR | AUC  | F1   | AUPR  | AUC   | F1   |
| SCMFDD       | 0.8377 | 0.8403 | 0.7765 | 0.8741 | 0.8750 | 0.8025 |
| EMP-SVD      | 0.9467 | 0.9444 | 0.8730 | 0.9786 | 0.9751 | 0.9163 |
| TS-SVD       | 0.9946 | 0.9933 | 0.9699 | 0.9994 | 0.9993 | 0.9947 |

Kitsiruwanwatt et al. [17]
The EMP-SVD model employed five meta-paths. Meta-path 1 was an adjacency matrix between drugs and diseases. Meta-path 2 described the path from drug to disease when passing a protein. Meta-path 3 described the path of drug-protein-drug-disease. Meta-path 4 described the drug-disease-drug-disease path. Meta-path 5 described the drug-disease-protein-disease path. We conducted experiments for every single meta-path and the ensemble meta-path and compared the performances of EMP-SVD when using the augmented and original drug-disease matrices. The performances of EMP-SVD with the original drug-protein and augmented drug-protein matrices are shown in Tables 4 and 5, respectively. The results showed that the performance of EMP-SVD with the augmented drug-protein matrix for all five meta-paths and the ensemble five meta-paths was superior to that of EMP-SVD with the original drug-protein matrix.

The EMP-SVD model was proposed based on five meta-paths, SVD techniques, and the reliable negative. The TS-SVD method was based on the common neighbors’ count matrix of drugs and diseases to achieve a topological similarity matrix, SVD, and a reliable negative, which was defined by k-step neighbors among drugs and diseases. We then investigated the performances of TS-SVD model with the augmented and original drug-protein matrices. These performances are presented in Table 6. The results showed that the TS-SVD with the augmented drug-protein matrix performed better than that with the original drug-protein matrix. Thus, our predictions of protein pairs sharing common drugs are efficient and can be used to improve the performance of existing drug repositioning models.

The augmented drug-protein matrix integrating the predictions of protein pairs sharing common drugs can discover the unrevealed associations of drug-associated proteins. The augmented drug-protein matrix enhances the performances of both EMP-SVD and TS-SVD. In EMP-SVD, the augmented drug-protein matrix influenced meta-path 2 and meta-path 3, which are composed of a direct path from the drug to protein. Tables 4 and 5 indicate that the augmented drug-protein matrix enhances the AUC values for meta-path 2 and meta-path 3 to 6.3% and 4.7%, respectively. Meanwhile, the augmented drug-protein matrix enhances the AUC value to 2.6%, 2.6%, 3.3%, and 3.3% for meta-paths 1, 4, and 5, respectively. In addition, EMP-SVD determined the reliable negative, which is the drug that cannot treat the disease, including that there are no common proteins between the drug and disease based on a heterogeneous network. Therefore, the augmented drug-protein matrix also impacts a reliable negative set for splitting a training

| Table 4. Performance of EMP-SVD when using the original drug-protein matrix. |
|-----------------|--------|--------|--------|--------|--------|--------|
| Meta-path | AUPR   | AUC    | PRE    | SEN    | ACC    | FI     |
| 1        | 0.937876 | 0.930704 | 0.870958 | 0.863608 | 0.865914 | 0.867194 |
| 2        | 0.897673 | 0.897781 | 0.850625 | 0.803789 | 0.820629 | 0.826455 |
| 3        | 0.919201 | 0.910094 | 0.848763 | 0.829973 | 0.836200 | 0.838955 |
| 4        | 0.938146 | 0.930872 | 0.867714 | 0.867984 | 0.869438 | 0.867831 |
| 5        | 0.931524 | 0.929587 | 0.888159 | 0.831903 | 0.853229 | 0.858980 |
| Ensemble | 0.946732 | 0.944475 | 0.893206 | 0.853809 | 0.869400 | 0.873000 |
set and a test set. Consequently, the augmented drug-protein matrix also has a minimal effect on the performance of meta-paths 1, 4, and 5.

Comparing the performance between the original and augmented drug-protein matrices based on the EMP-SVD and TS-SVD methods using the same set of drug-protein, disease-protein, and drug-disease associations, the performance scores of the TS-SVD method outperformed those of the EMP-SVD method when using the augmented drug-protein matrix. From the five-fold cross-validation, the prediction results from EMP-SVD and TS-SVD methods with prediction scores of more than 0.5 are shown in Supplementary Table S4 and S5, respectively.

### Table 5. Performance of EMP-SVD when using an augmented drug-protein matrix.

| Meta-path | AUPR   | AUC     | PRE    | SEN    | ACC    | F1     |
|-----------|--------|---------|--------|--------|--------|--------|
| 1         | 0.963804 | 0.955047 | 0.886046 | 0.917661 | 0.902514 | 0.901438 |
| 2         | 0.962469 | 0.954054 | 0.875309 | 0.919919 | 0.899171 | 0.897019 |
| 3         | 0.963386 | 0.953268 | 0.876241 | 0.910318 | 0.894400 | 0.892817 |
| 4         | 0.963871 | 0.955054 | 0.887318 | 0.916610 | 0.902686 | 0.901540 |
| 5         | 0.967072 | 0.960462 | 0.888840 | 0.916595 | 0.903629 | 0.902421 |
| Ensemble  | 0.978641 | 0.975175 | 0.917317 | 0.916026 | 0.915829 | 0.916323 |

### Table 6. Performance comparison of the TS-SVD method between the original and augmented drug-protein matrices.

| Performance | Prediction using original drug-protein matrix | Prediction using augmented drug-protein matrix |
|-------------|-----------------------------------------------|-----------------------------------------------|
| AUPR        | 0.994626                                      | 0.999454                                      |
| AUC         | 0.993322                                      | 0.999362                                      |
| PRE         | 0.961498                                      | 0.991664                                      |
| SEN         | 0.978513                                      | 0.997811                                      |
| ACC         | 0.969914                                      | 0.994571                                      |
| F1          | 0.969921                                      | 0.994727                                      |

Candidate drug-disease associations with the use of an augmented drug-protein matrix

New candidate drug-disease associations can be identified by selecting the false positives with high prediction scores from EMP-SVD and TS-SVD methods with the augmented matrix. The threshold for the candidate pairs was arbitrary to choose from the scores of range from 0.5 to 1.0. Therefore, with high confidence with scores of more than 0.9, there were about four drug-disease associations predicted by the EMP-SVD method (see Table 7) and only one drug-disease association found by the TS-SVD (see Table 8).
Table 7. The top 10 candidate drug-disease associations in the false positive group are predicted from the EMP-SVD method when using an augmented drug-protein matrix.

| DrugBank ID | Drug name   | DiseaseID | Disease name                      | Inference gene symbol | PubMedIDS            | Prediction score |
|-------------|-------------|-----------|-----------------------------------|-----------------------|----------------------|------------------|
| DB00177     | Valsartan   | MESH: D006943 | Hyperglycemia CCL2, CD163, COL3A1, IL6, NOS3, PTGS2 | 29035695, 20836762, 11696579, 14514642 | 0.9421875          |
| DB00158     | Folic Acid  | MESH: D020820 | Dyskinesias ALB, SLC22A2         | 10511253, 22525860    | 0.9164062          |
| DB01698     | Rutin       | MESH: D011041 | Poisoning ALB, SLC22A2         | 10511253, 22525860    | 0.9166667          |
| DB00945     | Acetylsalicylic acid | MESH: D004681 | Encephalomyelitis, Autoimmune, Experimental PPARA, SIRT1 | 17261635, 23547115 | 0.9053385          |
| DB00331     | Metformin   | MESH: D008527 | Medulloblastoma BRD2, CCNE1, CDK6, ESR2, IRS2, MYC, SKP2 | 24231268, 19270706, 19270706 | 0.8984375          |
| DB01211     | Clarithromycin | MESH: D003967 | Diarrhea IL1A | 11173893[9220047][9398876] 9570263[9855324, 9278552 | 0.8909598          |
| DB00947     | Fulvestrant | MESH: D000647 | Amnesia APP, CSF2, IL6, POMC | 12642396, 8877002, 8003924, 1989931, 2841920 | 0.8878478          |
| DB00338     | Omeprazole  | MESH: D012131 | Respiratory Insufficiency SLC23A1 | 11984597 | 0.88125          |
| DB00166     | Lipoic Acid | MESH: D058426 | Neointima AGT, MMP2              | 19258495[29609002, 17964422 | 0.880816          |
| DB09220     | Nicorandil  | MESH: D002375 | Catalepsy AGT                     | 1034924             | 0.8765625          |
However, to describe the relevance of a potential existing drug to treat another disease of the prediction results, the top 10 candidates of both methods were then validated with the present knowledge in the databases and literature as follows.

The top 10 novels of drug-disease pairs based on prediction scores from EMP-SVD and TS-SVD are shown in Tables 7 and 8, respectively. These pairs have not been reported as approved drug-disease pairs yet. Additionally, most of the novel drug-disease associations are found in the relationship based on the CTD database,\textsuperscript{40} which reports inferred associations by determining between chemical-gene interaction and disease-gene interaction. The gene that links chemicals and disease is called the inference gene. With the chemical-gene interaction and the disease-gene interaction on the CTD database, false positive pairs from EMP-SVD were further validated by their inference genes while all false positive pairs from TS-SVD were not found any inference genes involved with.

For the top prediction score of candidate drug-disease pairs from EMP-SVD using the augmented drug-protein matrix, valsartan (DB00177) is used to treat hypertension to lower the risk of cardiovascular events, such as strokes and myocardial infarctions; while hyperglycemia (MESH:D006943) is frequently associated with diabetes in which there is too much sugar in the blood due to a lack of insulin to transport glucose into the bloodstream.\textsuperscript{32} The CTD database reports that valsartan and hyperglycemia have an inferred association; moreover, inference genes are CCL2, CD163, COL3A1, IL6, NOS3, PTGS2.\textsuperscript{62,77–79} Furthermore, valsartan is assessed for the safety

\begin{table}[h]
\centering
\caption{The top 10 candidate drug-disease associations in false positive group are predicted from the TS-SVD method when using an augmented drug-protein matrix.}
\begin{tabular}{lllll}
\hline
DrugBankID & Drug name & DiseaseID & Disease name & Prediction score \\
\hline
DB01003 & Cromoglicic acid & MESH: D003555 & Cystinuria & 0.9140625 \\
DB00732 & Atracurium besylate & MESH: D056828 & Hereditary Angioedema Type III & 0.8984375 \\
DB14562 & Andexanet alfa & MESH: D014777 & Virus Diseases & 0.8671875 \\
DB11560 & Lesinurad & MESH: D012162 & Retinal Degeneration & 0.86328125 \\
DB00019 & Pegfilgrastim & MESH: D003555 & Cystinuria & 0.85546875 \\
DB04115 & Berberine & MESH: C563739 & Erythrokeratodermia Variabilis 3 & 0.8359375 \\
DB00348 & Nitisinone & MESH: D012162 & Retinal Degeneration & 0.8125 \\
DB04115 & Berberine & MESH: D020165 & Carbamoyl-Phosphate Synthase I Deficiency Disease & 0.80078125 \\
DB00023 & Asparaginase Escherichia coli & MESH: D012162 & Retinal Degeneration & 0.76171875 \\
DB01625 & Isopropamide & MESH: D012162 & Retinal Degeneration & 0.76171875 \\
\hline
\end{tabular}
\label{table:8}
\end{table}
and efficiency of diabetes in clinical testing (ClinicalTrials.gov Identifier: NCT00097786), the result indicates that valsartan has a relative reduction of 14% for patients with a glucose tolerance problem.\textsuperscript{80}

In addition, the CTD database presents that, rutin (DB01698) and poisoning (MESH: D011041),\textsuperscript{81,82} acetylsalicylic acid or aspirin (DB00945) and encephalomyelitis (MESH: D004681),\textsuperscript{83,84} metformin (DB00331) and medulloblastoma (MESH:D008527),\textsuperscript{85–88} clarithromycin (DB01211) and diarrhea (MESH:D003967),\textsuperscript{89} fulvestrant (DB00947) and amnesia (MESH:D000647),\textsuperscript{90–94} omeprazole (DB00338) and respiratory insufficiency (MESH:D012131),\textsuperscript{95} lipoic acid (DB00166) and neointima (MESH:D058426),\textsuperscript{96–98} and nicorandil (DB09220) and catalepsy (MESH:D002375)\textsuperscript{99} have reported as inferred associations between drugs and diseases, as shown in Table 7. Based on the CTD database, folic acid or vitamin B9 (DB00158) which is located in many supplements that are used to treat megaloablastic anemia\textsuperscript{32} therapeutic evidence is not found or an inferred association with dyskinesias (MESH:D020820). Dyskinesias is an unpredictable writhing movement of the face, arms, or legs. Dyskinesias indication is a side effect of certain Parkinson’s drugs such as levodopa.\textsuperscript{100} However, Folic Acid, vitamin B6, and vitamin B12 which are a supplement for Parkinson’s disease are evaluated for safety and efficiency of treatment in clinical testing (ClinicalTrials.gov Identifier: NCT00853879).

**Discussion**

This study exploits the various aspects of similarities among proteins to predict protein pairs that share common drugs and create an augmented drug-protein matrix to enhance the performance of drug repositioning approaches. The similarities among proteins in diversified biological meanings were categorized into four groups: PPI data, network, sequence alignment, and biological functions. The PPI data provide information on the relationships among protein pairs in physical and functional interactions. The network representing the associations among proteins indicates interactions between two proteins through the network structure and their neighborhood. Sequence alignment describes the structural and evolutionary similarities among protein pairs. The biological functions characterize the molecular interaction, reaction, and relation as well as the biological process among proteins.

For all 10 experiments, the forward selection algorithm yields the importance of features by ordering. The feature that appears first is the most important, while the one that turns up last is the least important, as shown in Supplementary Table S1. The local alignment feature is an outstanding feature for predicting protein pairs sharing common drugs. If the regions in the sequence of the two proteins are similar, then the common drugs can bind to these similar regions. The GO feature and common module feature are remarkable for most relevant features because they explain similar biological functions. The common drug might affect proteins in the same module of the complex system. The experimental feature is also one of the most relevant features for prediction. If the experimental data indicate directed interactions among proteins, they should operate together in some biological mechanisms. Then, the common drugs of these two proteins might relate to these two interacting proteins. The experimental results show that the fusion feature is irrelevant for our prediction, indicating that the score of an individual gene fused in other
species is not suitable for inferring common drugs between two proteins. From the feature selection algorithm, the crucial features are effect for predicting protein pairs sharing common drugs and yield a relatively high performance with an AUC of 0.939. Therefore, with the high prediction of predicting protein pairs, an augmented matrix for drugs and proteins was created to provide more relative information to infer new drug-disease associations.

To evaluate the performance of the use of an augmented drug-protein matrix for drug repositioning approaches, SCMFDD, EMP-SVD, and TS-SVD were implemented and compared. These three methods yielded better performance when using the augmented matrix. The SCMFDD method employed the drug feature similarities and disease semantic similarities for predicting a new drug-disease pair. The augmented matrix directly improved the protein target features for computing drug similarities. Thus, it resulted in better performance when using the augmented matrix. The EMP-SVD and TS-SVD were the methods based on meta-paths and topologies that directly applied a drug-protein matrix in their models. The EMP-SVD model was proposed based on five meta-paths, SVD techniques, and the reliable negative. Recently, a drug-repositioning method with topological similarity and singular value decomposition (TS-SVD), an improved version of EMP-SVD, was developed. The TS-SVD method was based on the common neighbors’ count matrix of drugs and diseases to achieve a topological similarity matrix, SVD, and a reliable negative, which was defined by k-step neighbors among drugs and diseases. We then investigated the performances of the TS-SVD model with the original and augmented drug-protein matrices. The performances were improved from 0.9946 to 0.9995 using the augmented matrices. The performance scores of the TS-SVD are quite high because the TS-SVD performed a filtering process of the reliable negative sample before fitting the model. We investigated the impact of the filtering process and compared the performances with filtering and without filtering reliable negative samples based on drug-protein associations via DrugBank database are shown in the Supplementary Table S6. The performances were decreased when generating a predictive model without selecting reliable negative samples. This result showed that filtering the reliable negative data based on k-step neighbors among drugs and diseases impacts predicting the drug repositioning approach. However, this filtering process might be considered to bias by separating positive and real negative samples to refine the data before generating a predictive model. Nevertheless, the augmented drug-protein matrix could improve the performance showing that the TS-SVD with the augmented drug-protein matrix performed better than that with the original drug-protein matrix. Thus, our predictions of protein pairs sharing common drugs are efficient and can be used to improve the performance of existing drug repositioning models.

Our study reveals 10 promising protein pairs sharing the same drugs, whose corresponding drugs show high similarity in terms of the chemical structure. However, these protein pairs have not been reported as directed evidence of sharing common drugs in the database. Therefore, these potential pairs are compelling for further studies of drug development and design. One limitation of this study is the insufficient data available for approved drugs related to protein pairs, generating noisy data in the negative set.
Conclusions

This study proposes the use of a random forest model for predicting protein pairs that share common drugs. The features of the model were based on the relationships among proteins in various biological aspects. This study could be useful in enhancing the opportunity to discover missing drug-associated proteins. The forward selection technique provides a set of crucial features for prediction. The results obtained from the feature selection technique revealed crucial features, such as local alignment, GO, experiments, common modules, text mining, global alignment, co-expression, common neighbors, conserved neighborhood, closer proteins, databases, and co-occurrence. The model yielded high performance, with an AUC exceeding 93%.

Moreover, this study suggested 10 novel potential protein pairs predicted to share common drugs. These protein pairs have not yet been known to share the same drug in DrugBank. They achieved a very high chemical structure similarity score. Therefore, these 10 protein pairs are very interesting for further investigation.

Furthermore, this study proposes an augmented drug-protein matrix based on the predictions of protein pairs sharing the same drugs. The augmented matrix can enhance the performance of existing drug repositioning techniques. Hence, the predictions of protein pairs sharing the same drugs can be used for drug repositioning.

Author contributions

Conceptualization, S.K., A.S. and K.P.; methodology, S.K., A.S. and K.P.; funding acquisition, A.S.; formal analysis, S.K.; validation, S.K., A.S. and K.P.; writing—original draft preparation, S.K.; writing—review and editing, S.K., A.S. and K.P.; supervision K.P. and A.S.

All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data sharing is not applicable to this article.

Acknowledgments

We would like to acknowledge National e-Science Infrastructure Consortium (http://www.e-science.in.th) (accessed on 11/10/2020) for kindly supporting the high-performance computing resources.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the King Mongkut’s University of Technology North Bangkok, (grant number KMUTNB-64-KNOW-21).

ORCID iDs
Apichat Suratanee https://orcid.org/0000-0002-3439-6619
Kitiporn Plaimas https://orcid.org/0000-0002-0416-8316

Supplemental material
Supplemental material for this article is available online

References
1. Gottlieb A, Stein GY, Ruppin E, et al. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol* 2011; 7: 1–9.
2. Blockeel H, Kersting K, Nijsen S, et al. Machine Learning and Knowledge Discovery in Databases. 2013.
3. Ashburn TT and Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discovery* 2004; 3: 673–683.
4. Sleigh SH and Barton CL. Repurposing strategies for therapeutics. *Pharmaceut Med* 2010; 24: 151–159.
5. Roses AD. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat Rev Drug Discovery* 2008; 7: 807–817.
6. Shim JS and Liu JO. Recent advances in drug repositioning for the discovery of new cancer drugs. *Int J Biol Sci* 2014; 10: 654–663.
7. Khalid Z and Sezerman OU. Computational drug repurposing to predict approved and novel drug-disease associations. *J Mol Graphics Modell* 2018; 85: 91–96.
8. Hodos RA, Kidd BA, Shameer K, et al. In silico methods for drug repurposing and pharmacology. Wiley interdisciplinary reviews. *Systems Biology and Medicine* 2016; 8: 186–210.
9. Zhang W, Yue X, Lin W, et al. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics* 2018; 19: 233. –.
10. Tabei Y, Kotera M, Sawada R, et al. Network-based characterization of drug-protein interaction signatures with a space-efficient approach. *BMC Syst Biol* 2019; 13: 39.
11. Hu S, Zhang C, Chen P, et al. Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks. *BMC bioinformatics* 2019; 20: 1–12.
12. Hu S-S, Chen P, Wang B, et al. Protein binding hot spots prediction from sequence only by a new ensemble learning method. *Amino Acids* 2017; 49: 1773–1785.
13. Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008; 24: i232–i240.
14. Hodos RA and Kidd BA. Computational Approaches to Drug Repurposing and Pharmacology. HHS Public Access. 2017:1–46.
15. Cheng F, Liu C, Jiang J, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012; 8: e1002503.
16. Bleakley K and Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* 2009; 25: 2397–2403.
17. Ding H, Mamitsuka H and Zhu S. Similarity-based machine learning methods for predicting drug-target interactions: a brief review. *Briefings in Bioinformatics* 2013; 15: 734–747.
18. Suratanee A and Plaimas K. DDA: a novel network-based scoring method to identify disease-disease associations. *Bioinform Biol Insights* 2015; 9: 175–186.
19. Suratanee A, Buaboocha T and Plaimas K. Prediction of human-plasmodium vivax protein associations from heterogeneous network structures based on machine-learning approach. *Bioinform Biol Insights* 2021; 15: 11779322211013350.
20. Suratanee A and Plaimas K. Identification of inflammatory bowel disease-related proteins using a reverse k-nearest neighbor search. *J Bioinform Comput Biol* 2014; 12: 1450017.
21. Suratanee A and Plaimas K. Reverse nearest neighbor search on a protein-protein interaction network to infer protein-disease associations. *Bioinform Biol Insights* 2017; 11: 1177932217720405.
22. Suratanee A and Plaimas K. Network-based association analysis to infer new disease-gene relationships using large-scale protein interactions. *PLoS One* 2018; 13: e0199435.
23. Suratanee A and Plaimas K. Heterogeneous network model to identify potential associations between plasmodium vivax and human proteins. *Int J Mol Sci* 2020; 21: 1310.
24. Suratanee A and Plaimas K. Hybrid deep learning based on a heterogeneous network profile for functional annotations of plasmodium falciparum genes. *Int J Mol Sci* 2021; 22: 10019.
25. Janyasupab P, Suratanee A and Plaimas K. Network diffusion with centrality measures to identify disease-related genes. *Math Biosci Eng* 2021; 18: 2909–2929.
26. Kawichai T, Suratanee A and Plaimas K. Meta-Path based gene ontology profiles for predicting drug-disease associations. *IEEE Access* 2021; 9: 41809–41820.
27. Kitsiranuwat S, Suratanee A and Plaimas K. Multi-Data aspects of protein similarity with a learning technique to identify drug-disease associations. *Applied Sciences* 2021; 11: 2914.
28. Wu G, Liu J and Yue X. Prediction of drug-disease associations based on ensemble meta paths and singular value decomposition. *BMC Bioinformatics* 2019; 20: 1–13.
29. Wu G and Liu J. Predicting Drug-Disease Treatment Associations Based on Topological Similarity and Singular Value Decomposition. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2019. p. 153–158.
30. Szklarczyk D, Gable AL, Lyon D, et al. STRING V11: protein – protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019; 47: 607–613.
31. Bateman A. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019; 47: D506–D515.
32. Wishart DS, Feunang YD, Guo AC, et al. Drugbank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018; 46: D1074–D1082.
33. Xiao N, Cao DS, Zhu MF, et al. Protr/ProtrWeb: r package and web server for generating various numerical representation schemes of protein sequences. *Bioinformatics* 2015; 31: 1857–1859.
34. Yi Y, Fang Y, Wu K, et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2020; 19: 3316–3332.
35. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019; 28: 1947–1951.
36. Kanehisa M, Sato Y, Furumichi M, et al. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019; 47: D590–D595.
37. Huntley RP, Sawford T, Mutowo-Meullenet P, et al. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res* 2015; 43: D1057–D1063.
38. Wishart DS, Knox C, Guo AC, et al. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006; 34: 668–672.
39. Guha R. Chemical informatics functionality in R. *J Stat Softw* 2007; 18: 1–16.
40. Davis AP, Grondin CJ, Johnson RJ, et al. The comparative toxicogenomics database: update 2019. *Nucleic Acids Res* 2019; 47: D948–D954.
41. Pi J, Ram JM, Sa J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020; 48: 845–855.
42. Mount D. *Uoa, tucson. Bioinformatics: sequence and genome analysis.* Second Edition. New York: Cold Spring Harbor Laboratory Press, 2004.
43. Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis* 2020; 71: 762–768.
44. Liu Y, Yang Y, Zhang C, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci* 2020; 63: 364–374.
45. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497–506.
46. Kane JM, Cohen M, Zhao J, et al. Efficacy and safety of asenapine in a placebo- and haloperidol-controlled trial in patients with acute exacerbation of schizophrenia. *J Clin Psychopharmacol* 2010; 30: 106–115.
47. Habre W, Wilson D and Johnson CM. Extrapyramidal side-effects from droperidol mixed with morphine for patient-controlled analgesia in two children. *Paediatr Anaesth* 1999; 9: 362–364.
48. Maslov LN and Lishmanov Iu B. Antiarrhythmic properties of opioid receptor agonists. *Eksp Klin Farmakol* 2006; 69: 69–79.
49. Langheinrich U, Vacun G and Wagner T. Zebrafish embryos express an orthologue of HERG and are sensitive toward a range of QT-prolonging drugs inducing severe arrhythmia. *Toxicol Appl Pharmacol* 2003; 193: 370–382.
50. Wolpert C, Schimpf R, Giustetto C, et al. Further insights into the effect of quinidine in short QT syndrome caused by a mutation in HERG. *J Cardiovasc Electrophysiol* 2005; 16: 54–58.
51. Hassel D, Scholz EP, Trano N, et al. Deficient zebrafish ether-à-go-go-related gene channel gating causes short-QT syndrome in zebrafish reggae mutants. *Circulation* 2008; 117: 866–875.
52. Yoshihara T, Ishigaki S, Yamamoto M, et al. Differential expression of inflammation- and apoptosis-related genes in spinal cords of a mutant SOD1 transgenic mouse model of familial amyotrophic lateral sclerosis. *J Neurochem* 2002; 80: 158–167.
53. Li M, Ona VO, Guégan C, et al. Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model. *Science* 2000; 288: 335–339.
54. Lee M, Hyun D, Jenner P, et al. Effect of overexpression of wild-type and mutant Cu/Zn-superoxide dismutases on oxidative damage and antioxidant defences: relevance to Down’s Syndrome and familial amyotrophic lateral sclerosis. *J Neurochem* 2001; 76: 957–965.
55. Takehisa Y, Ujike H, Ishizu H, et al. Familial amyotrophic lateral sclerosis with a novel Leu126Ser mutation in the copper/zinc superoxide dismutase gene showing mild clinical features and Lewy body-like hyaline inclusions. *Arch Neurol* 2001; 58: 736–740.
56. Chioza BA, Ujfalussy A, Csiszar K, et al. Mutations in the lysyl oxidase gene are not associated with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2001; 2: 93–97.
57. Giess R, Holtmann B, Braga M, et al. Early onset of severe familial amyotrophic lateral sclerosis with a SOD-1 mutation: potential impact of CNTF as a candidate modifier gene. *Am J Hum Genet* 2002; 70: 1277–1286.
58. Wang J, Xu G, Gonzales V, et al. Fibrillar inclusions and motor neuron degeneration in transgenic mice expressing superoxide dismutase 1 with a disrupted copper-binding site. *Neurobiol Dis* 2002; 10: 128–138.
59. de Beus MD, Chung J and Colón W. Modification of cysteine 111 in Cu/Zn superoxide dismutase results in altered spectroscopic and biophysical properties. *Protein Sci* 2004; 13: 1347–1355.

60. Rembach A, Turner BJ, Bruce S, et al. Antisense peptide nucleic acid targeting GluR3 delays disease onset and progression in the SOD1 G93A mouse model of familial ALS. *J Neurosci Res* 2004; 77: 573–582.

61. Fujiwara N, Miyamoto Y, Ogashahara K, et al. Different immunoreactivity against monoclonal antibodies between wild-type and mutant copper/zinc superoxide dismutase linked to amyotrophic lateral sclerosis. *J Biol Chem* 2005; 280: 5061–5070.

62. Banci L, Bertini I, D’Amelio N, et al. Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. *J Biol Chem* 2005; 280: 35815–35821.

63. Winkler DD, Schuermann JP, Cao X, et al. Structural and biophysical properties of the pathogenic SOD1 variant H46R/H48Q. *Biochemistry* 2009; 48: 3436–3447.

64. Milardi D, Pappalardo M, Grasso DM, et al. Unveiling the unfolding pathway of FALS associated G37R SOD1 mutant: a computational study. *Mol Biosyst* 2010; 6: 1032–1039.

65. Arciello M, Capo CR, D’Annibale S, et al. Copper depletion increases the mitochondrial-associated SOD1 in neuronal cells. *Biometals* 2011; 24: 269–278.

66. Wright GS, Antonyuk SV, Kershaw NM, et al. Ligand binding and aggregation of pathogenic SOD1. *Nat Commun* 2013; 4: 1758.

67. Tokuda E, Okawa E, Watanabe S, et al. Overexpression of metallothionein-I, a copper-regulating protein, attenuates intracellular copper dyshomeostasis and extends lifespan in a mouse model of amyotrophic lateral sclerosis caused by mutant superoxide dismutase-1. *Hum Mol Genet* 2014; 23: 1271–1285.

68. Lin PY, Simon SM, Koh WK, et al. Heat shock factor 1 overexpression protects against exposure of hydrophobic residues on mutant SOD1 and early mortality in a mouse model of amyotrophic lateral sclerosis. *Mol Neurodegener* 2013; 8: 43.

69. Furukawa Y, Anzai I, Akiyama S, et al. Conformational disorder of the most immature Cu,Zn-superoxide dismutase leading to amyotrophic lateral sclerosis. *J Biol Chem* 2016; 291: 4144–4155.

70. Gurney ME, Cutting FB, Zhai P, et al. Benefit of vitamin E, riluzole, and gabapentin in a transgenetic model of familial amyotrophic lateral sclerosis. *Ann Neurol* 1996; 39: 147–157.

71. Chen L, Liu HG, Liu W, et al. Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia. *Zhonghua Jie He Hu Xi Za Zhi* 2020; 43: E005.

72. Yang Q, Xie DR, Jiang ZM, et al. Efficacy and adverse effects of transdermal fentanyl and sustained-release oral morphine in treating moderate-severe cancer pain in Chinese population: a systematic review and meta-analysis. *J Exp Clin Cancer Res* 2010; 29: 67.

73. Goundrey J. Vertigo after epidural morphine. *Can J Anaesth* 1990; 37: 804–805.

74. Clausen SF, Schneider D and Patil NP. The treatment of minocycline-induced brainstem vertigo by the combined administration of piracetam and ergotoxin. *Acta Otolaryngol Suppl* 1989; 468: 171–174.

75. Hu H, Luo ML, Du XL, et al. Up-regulated manganese superoxide dismutase expression increases apoptosis resistance in human esophageal squamous cell carcinomas. *Chin Med J (Engl)* 2007; 120: 2092–2098.

76. Zhang J, Wang K, Zhang J, et al. Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma. *J Proteome Res* 2011; 10: 2863–2872.

77. Gaikwad AB, Gupta J and Tikoo K. Epigenetic changes and alteration of Fbn1 and Col3A1 gene expression under hyperglycaemic and hyperinsulinaemic conditions. *Biochem J* 2010; 432: 333–341.
78. Du XL, Edelstein D, Dimmeler S, et al. Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site. J Clin Invest 2001; 108: 1341–1348.

79. Kiritoshi S, Nishikawa T, Sonoda K, et al. Reactive oxygen species from mitochondria induce cyclooxygenase-2 gene expression in human mesangial cells: potential role in diabetic nephropathy. Diabetes 2003; 52: 2570–2577.

80. McMurray JJ, Holman RR, Haffner SM, et al. Effect of valsartan on the incidence of diabetes and cardiovascular events. N Engl J Med 2010; 362: 1477–1490.

81. Rappaport SM and Yeowell-O’Connell K. Protein adducts as dosimeters of human exposure to styrene, styrene-7,8-oxide, and benzene. Toxicol Lett 1999; 108: 117–126.

82. Zhang J and Zhou W. Ameliorative effects of SLC22A2 gene polymorphism 808 G/T and cimetidine on cisplatin-induced nephrotoxicity in Chinese cancer patients. Food Chem Toxicol 2012; 50: 2289–2293.

83. Dunn SE, Ousman SS, Sobel RA, et al. Peroxisome proliferator-activated receptor (PPAR) alpha expression in T cells mediates gender differences in development of T cell-mediated autoimmune. J Exp Med 2007; 204: 321–330.

84. Nimmagadda VK, Bever CT, Vattikunta NR, et al. Overexpression of SIRT1 protein in neurons protects against experimental autoimmune encephalomyelitis through activation of multiple SIRT1 targets. J Immunol 2013; 190: 4595–4607.

85. Henssen A, Thor T, Odersky A, et al. BET Bromodomain protein inhibition is a therapeutic option for medulloblastoma. Oncotarget 2013; 4: 2080–2095.

86. Northcott PA, Nakahara Y, Wu X, et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. Nat Genet 2009; 41: 465–472.

87. Whiteway SL, Harris PS, Venkataraman S, et al. Inhibition of cyclin-dependent kinase 6 suppresses cell proliferation and enhances radiation sensitivity in medulloblastoma cells. J Neurooncol 2013; 111: 127–132.

88. Mancuso M, Leonardi S, Ceccarelli M, et al. Protective role of 17 β-estradiol on medulloblastoma development in patched 1 heterozygous mice. Int J Cancer 2010; 127: 2749–2757.

89. Huang SL, Takagi Y, Wasa M, et al. Induction of nitric oxide synthase in rat intestine by interleukin-1alpha may explain diarrhea associated with zinc deficiency. J Nutr 1997; 127: 1729–1736.

90. Wang HH, Chien JW, Chou YC, et al. Anti-amnesic effect of dimemorfan in mice. Br J Pharmacol 2003; 138: 941–949.

91. Bianchi M, Sacerdote P and Panerai AE. Peripherally administered GM-CSF interferes with scopolamine-induced amnesia in mice: involvement of interleukin-1. Brain Res 1996; 729: 285–288.

92. Bianchi M and Panerai AE. Peripherally administered IL-1 alpha interferes with scopolamine-induced amnesia in mice. Brain Res Cogn Brain Res 1993; 1: 257–259.

93. Bianchi M, Ferrario P, Clavenna A, et al. Interleukin-6 affects scopolamine-induced amnesia, but not brain amino acid levels in mice. Neuroreport 1997; 8: 1775–1778.

94. Dalmaz C, Godoy MG and Izquierdo I. Post-training and pretest effects of adrenocorticotropic hormone on retention: the influence of the hour of the day, the training-test interval, and pretest naloxone administration. Behav Neural Biol 1988; 49: 406–411.

95. Sotiriou S, Gispert S, Cheng J, et al. Ascorbic-acid transporter Scl23a1 is essential for vitamin C transport into the brain and for perinatal survival. Nat Med 2002; 8: 514–517.

96. Lacchini S, Heimann AS, Evangelista PS, et al. Cuff-induced vascular intima thickening is influenced by titration of the Ace gene in mice. Physiol Genomics 2009; 37: 225–230.

97. Lee DY, Won KJ, Lee KP, et al. Angiotensin II facilitates neointimal formation by increasing vascular smooth muscle cell migration: involvement of APE/Ref-1-mediated overexpression of sphingosine-1-phosphate receptor 1. Toxicol Appl Pharmacol 2018; 347: 45–53.
98. Wu TC, Chen YH, Leu HB, et al. Carvedilol, a pharmacological antioxidant, inhibits neointimal matrix metalloproteinase-2 and −9 in experimental atherosclerosis. *Free Radic Biol Med* 2007; 43: 1508–1522.

99. Braszko J and Wiśniewski K. Effect of angiotensin on the central action of dopamine. *Pol J Pharmacol Pharm* 1976; 28: 667–672.

100. Marconi R, Lefebvre-Caparros D, Bonnet A-M, et al. Levodopa-induced dyskinesias in Parkinson’s Disease phenomenology and pathophysiology. *Mov Disord* 1994; 9: 2–12.