

Research Article

Deep Learning-Assisted Repurposing of Plant Compounds for Treating Vascular Calcification: An In Silico Study with Experimental Validation

Chia-Ter Chao ^{1,2,3}, You-Tien Tsai,³ Wen-Ting Lee,⁴ Hsiang-Yuan Yeh ⁴
and Chih-Kang Chiang ²

¹Nephrology Division, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

²Graduate Institute of Toxicology, National Taiwan University College of Medicine, Taipei, Taiwan

³Nephrology Division, Department of Internal Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

⁴School of Big Data Management, Soochow University, Taipei, Taiwan

Correspondence should be addressed to Hsiang-Yuan Yeh; hyyeh.richard@gmail.com

Received 6 July 2021; Revised 24 October 2021; Accepted 13 November 2021; Published 5 January 2022

Academic Editor: Vladimir Jakovljevic

Copyright © 2022 Chia-Ter Chao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Vascular calcification (VC) constitutes subclinical vascular burden and increases cardiovascular mortality. Effective therapeutics for VC remains to be procured. We aimed to use a deep learning-based strategy to screen and uncover plant compounds that potentially can be repurposed for managing VC. **Methods.** We integrated drugome, interactome, and diseasome information from Comparative Toxicogenomic Database (CTD), DrugBank, PubChem, Gene Ontology (GO), and BioGrid to analyze drug-disease associations. A deep representation learning was done using a high-level description of the local network architecture and features of the entities, followed by learning the global embeddings of nodes derived from a heterogeneous network using the graph neural network architecture and a random forest classifier established for prediction. Predicted results were tested in an *in vitro* VC model for validity based on the probability scores. **Results.** We collected 6,790 compounds with available Simplified Molecular-Input Line-Entry System (SMILES) data, 11,958 GO terms, 7,238 diseases, and 25,482 proteins, followed by local embedding vectors using an end-to-end transformer network and a node2vec algorithm and global embedding vectors learned from heterogeneous network via the graph neural network. Our algorithm conferred a good distinction between potential compounds, presenting as higher prediction scores for the compound categories with a higher potential but lower scores for other categories. Probability score-dependent selection revealed that antioxidants such as sulforaphane and daidzein were potentially effective compounds against VC, while catechin had low probability. All three compounds were validated *in vitro*. **Conclusions.** Our findings exemplify the utility of deep learning in identifying promising VC-treating plant compounds. Our model can be a quick and comprehensive computational screening tool to assist in the early drug discovery process.

1. Introduction

The ectopic deposition of calcium apatite crystals in vascular walls, or vascular calcification (VC), emerges as a silent killer for populations with chronic kidney disease (CKD), diabetes mellitus (DM), and those of advanced age [1]. VC can be detected in as early as the fourth decade of life and becomes

nearly universal in those older than 65 years, progressing from the abdominal aorta to the ascending aorta at a later time [2]. Meta-analyses showed that the prevalence of coronary or aortic VC reached as high as 65% to 70% [3, 4]. VC leads to vascular stiffening, reducing major vessel compliance and increasing cardiac afterload, culminating in myocardial hypertrophy, heart failure, and cardiovascular

mortality. Patients with CKD are at a particularly high risk [5], and they frequently develop occult vascular pathologies especially VC. The perverse influences conferred by VC therefore warrants complete understandings of its pathophysiology and fuel searches for potential candidates to treat VC [6, 7].

Processes responsible for the development and aggravation of VC include not only passive calcium deposition but also active osteoid secretion by resident cells (especially vascular smooth muscle cells (VSMCs)) [8]. The adverse micro-environment, including dysregulated mineral metabolism-related hormones (fibroblast growth factor-23 (FGF-23), parathyroid hormone, and vitamin D), divalent ion imbalances, and/or the downregulation of mineralization inhibitors prompts the *trans*-differentiation of VSMCs into osteoblast-like cells with a secretory phenotype [9]. Uremic toxins [10] and epigenetic machineries (microRNAs and methylation abnormalities) [11] also participate in the progressive courses of VC. The complex pathogenesis of VC creates an unformidable challenge for identifying effective therapeutics. Despite the uncovering of potential candidates, very few enter clinical trials and none is recommended by international guidelines [12]. Therapeutic nihilism therefore prevails regarding how to treat VC [13].

The advancement in computational biology and high-throughput platform for target/molecule screening greatly facilitates our understanding of disease pathophysiology and novel treatment identification, including VC [14]. Compounds of either synthetic origin or natural molecules have been shown to be a rich reservoir for procuring potential therapeutic candidates [6]. Druggable targets inherent to the pathogenic landscape of VC can also be discovered through similar strategies. Using a transcriptomic approach followed by experimental validation, we previously showed that astaxanthin might beneficially influence VC severity through its action toward superoxide dismutase 2 (SOD2) [15]. Machine learning, a new strategy and expanding research discipline that trains computers to learn from massive data, emerges as a promising tool to help mine patterns or linkages [16]. The application of machine learning to the field of medicine has fueled and accelerated the progress of precision medicine. Existing literature concerning the use of machine learning in VC studies unanimously focuses on image interpretations and cardiovascular risk stratification, but rarely on plausible therapeutic candidate selection [17]. Furthermore, the process of new drug discovery is costly and time consuming with low success rates. We need more economically efficient ways such as existing drug repurposing, which aims at maximizing the potential usage of existing drugs [18]. As the effective clinical treatment for VC remains a largely undercharted field, we aimed to use a deep learning-based strategy to screen and identify existing compounds capable of being repurposed for managing this dreadful disorder.

2. Material and Methods

2.1. Overall Workflow of Database Integration and Deep Learning. We used the representation learning strategy,

which is aimed at providing a meaningful embedding vector for each node in heterogeneous graphs in order to facilitate the binary classification application of the drug-disease association [19, 20]. A gross overview of our workflow is illustrated in Figure 1. In step 1, we combined the drugome, interactome, and disease information from different biomedical database to analyze drug-disease associations. In step 2, we learned a high-level description of the local network architecture and features of the entities using a deep representation learning. This step encoded the content embedding for each node; for instance, the drug might have its own chemical structure for functionality and interactions with its confirmed or predicted target proteins as a subnetwork. In step 3, we learned the global embeddings of the nodes derived from the heterogeneous network using the graph neural network architecture, which was treated as an input into a classification task. This step aggregated the content embeddings from different neighbors and then combined them together by considering different impacts from different nodes. In step 4, we applied a random forest classifier to predict novel associations based on the embedding vectors of the selected drugs and the disease(s) of interest.

2.2. Integrating Heterogeneous Biological Information between Drugs and Diseases. The fields of “network medicine” and “system biology” are gaining tractions in the recent years. An emerging application of such strategy is to uncover novel associations based on a diverse spectrum of relationships between drugs and diseases [21]. We gathered networks connecting drugs, interactomes, and phenotypes from different biomedical databases and also incorporated heterogeneous information for considering vast arrays of attributes or contents associated with each node in Figure 2(a). The descriptions of each drug were gathered from 2D chemical structures in PubChem, drug-target associations in Comparative Toxicogenomics Database (CTD) [22], DrugBank databases [23], and functional annotations in Gene Ontology (GO). The CTD database provides the disease phenotype-genotype associations and disease phenotype-drug associations. We curated 9,726 drugs, 25,482 target proteins, and 7,238 diseases, while interactions among proteins of human origin were extracted from Bio-Grid containing 15,352 unique proteins and 281,862 interactions. We used the enriched GO annotations, which were statistically significant GO terms with a p value < 0.01 for compounds recorded in the CTD database.

2.3. Structural Characteristic Embeddings of Chemical Drug Using the End-to-End Transformer. Cheminformatics is aimed at integrating chemistry with information science techniques based on interactions, structural characteristics, and functional properties [24]. We extracted the Canonical Simplified Molecular-Input Line-Entry System (SMILES) format of the drugs from the PubChem database. The structure of the end-to-end transformer deep neural network is shown in Figure 2(b). We applied the pretrained end-to-end transformer to 83,000,000 SMILES collected from PubChem to obtain the structural characteristic embedding vectors of each drug using the end-to-end transformer deep

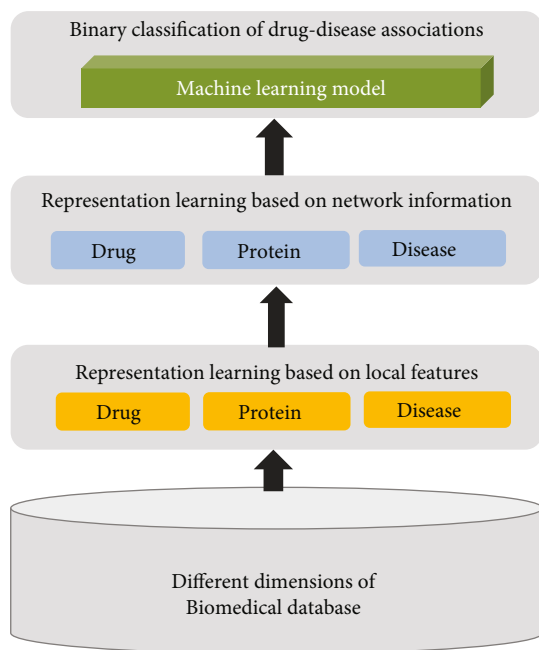


FIGURE 1: A schematic overview of our workflow.

neural network based on their frequency and sequential order of each SMILES character. An encoder layer with self-attention operation mapped the SMILES sequence into the latent space based on its relationship with other characters [25]. A decoder layer had similar structures to encoder layers, and the output of the final decoder layer was the same as the input sequence (Figure 2(b)). The output of the final encoder layer provided the structural characteristic embeddings of a chemical compound, representing the abstract features for describing the structural characteristics of a chemical drug. We further applied the *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm to visualize complex high-dimensional content embedding vectors in a two-dimensional space.

2.4. Functional Characteristic Embeddings of Chemical Compounds. To study functional properties, we gathered drug-GO term relationships as bipartite graphs and harnessed *node2vec*, an algorithmic framework for learning the functional characteristics embeddings of the chemical drug in bipartite networks [26]. We first used a biased random walk algorithm to sample sequential paths with different lengths to capture indirect relationships, or high-order dependencies in the bipartite graph. We treated the sequential paths as sentences in natural language and applied the Skip-gram model to generate low-dimensional and real vector space representing the functional characteristics embeddings of each chemical compound. Importantly, compounds with similar functional properties could be mapped closer to each other on the latent space.

2.5. Representative Learning of Nodes in Heterogeneous Network Using Graph Neural Networks. Besides the chemical structural and functional characteristic embedding vectors of drugs, the embedding vectors of the diseases and proteins

are also learned from disease-gene and drug-target protein interaction networks using the *node2vec* algorithm. After obtaining initial embedding vectors, we summarized a graph-level latent vector considering the heterogeneous structural information via graph deep neural network architectures (GNN) as in Figure 2(c) [27]. We analyzed the latent vectors with the geometric relationships which reflected the topological relationships between nodes. GNN iteratively updated the initial embedding vectors learned from the features of the nodes by propagating the information from their neighbors based on the attention mechanism [27, 28]. An aggregation function took the current embedding vectors of the neighbors of a node and created comprehensive node embedding that maximized effects from the neighborhoods of nodes with different weights as global embedding vectors.

2.6. Random Forest Classifier for Binary Classification between Drugs and Diseases. We combined the drug and disease global embedding vectors together as the input of the random forest classifier for predicting associations between drugs and diseases (Figure 3). We adopted known drug-disease associations in CTD as the gold standard and applied fivefold cross-validation to evaluate the performances of our constructed model. We randomly split all known drug-disease associations into five equal-sized subsets, combined four subsets as the training set, and treated the other subset as the testing set. We estimated the average of performances ranging from 0.65 to 0.76 using the area under the receiver operating characteristics curve (AUC) value on 5-fold cross-validations. Plant compounds with a probability score predicted by our model higher than 0.5 were selected for testing in experimental models of VC, as detailed below. We also selected one candidate with a low probability score for validation.

2.7. Comparing the Embedding Vectors Based on Different Information: The Similarity-Based Strategy. We used a similarity-based strategy to compare the embedding vectors based on different information supported. Similarity-based compound virtual screening using the sparse and binary fingerprint or functional annotation is one of the fundamental methods of preclinical drug discovery. This strategy harnessed computational approaches to predict properties and potential biologic effects of compounds. For comparison purpose, we used the aforementioned dense embedding vectors based on structural (SMILES-only) using end-to-end transformers and chemical functional activities (GO annotations-only) derived from the *node2vec* algorithm. We chose the cosine similarity measurement to formulate the diversity via the angular between different vectors in the embedding, and the score lied within the interval [0, 1], where the value 0 or 1 indicated an extremely dissimilar or extremely similar probability to the reference standards, respectively. Particularly, the diversity mainly provided a representative model with more complement information for the machine learning process [29]. We considered two structural embedding vectors to be significantly similar if the cosine similarity

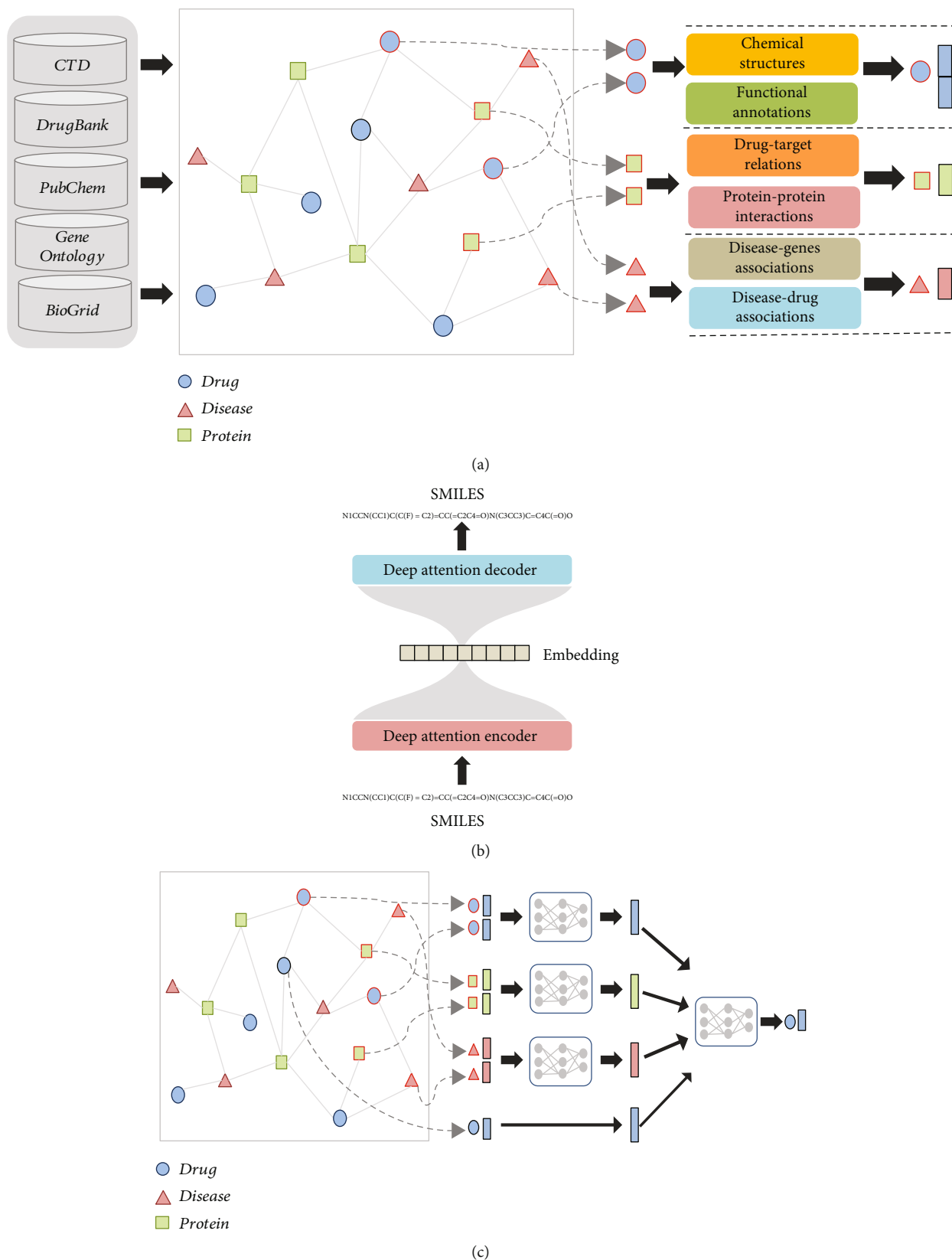


FIGURE 2: The steps of representation learning in our model. (a) Integrating different dimensions of biomedical database into a heterogeneous network and representation learning with their local content features. (b) The end-to-end transformer for chemical SMILES representation learning. (c) Aggregating the local content features to global features through the heterogeneous network using the graph neural network. CTD: Comparative Toxicogenomic Database; SMILES: Simplified Molecular-Input Line-Entry System.

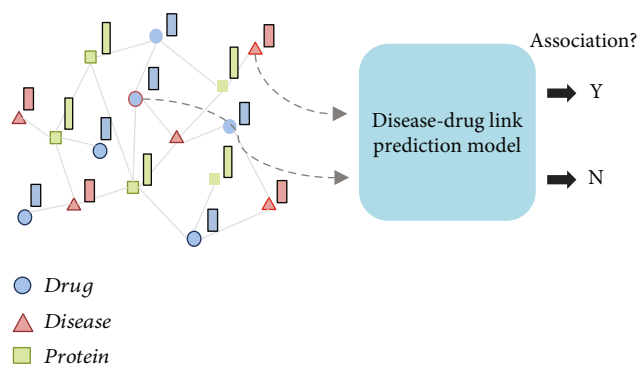


FIGURE 3: Binary classification of drug-disease associations based on the embedding vectors of the selected compounds and the disease(s) of interest.

scores were above 0.85 between them, according to prior reports [30].

2.8. Testing Compounds for Therapeutic Use against the Predicted Phenotype, VC. We harnessed an experimental model of VC established previously [10, 11, 15] for testing identified compounds. Briefly, a VSMC cell line, A7r5, between 3 and 6 passages was obtained from American Type Culture Collection, plated onto 24- and 6-well plates, and cultured with Dulbecco’s modified Eagle’s medium (DMEM) supplemented with fetal bovine serum (FBS), streptomycin/penicillin in 5% ambient CO₂ under 37°C. For the induction of biomineralization, culture media of cells reaching 80% confluence were switched to conditioned ones containing 2.5 to 3 mM inorganic phosphate and adjuvants, resembling the calcification-prone milieu observed in patients with CKD or those of advanced age. Condition media were exchanged every 2 days. After 7 days of treatment, cells would exhibit morphological alterations from being spindle-shaped to assuming a flattened and distended contour, accompanied by prominent extracellular calcium deposition. We qualitatively examined the extent of VC using Alizarin Red staining, with calcifying regions showing reddish areas in nodular aggregates.

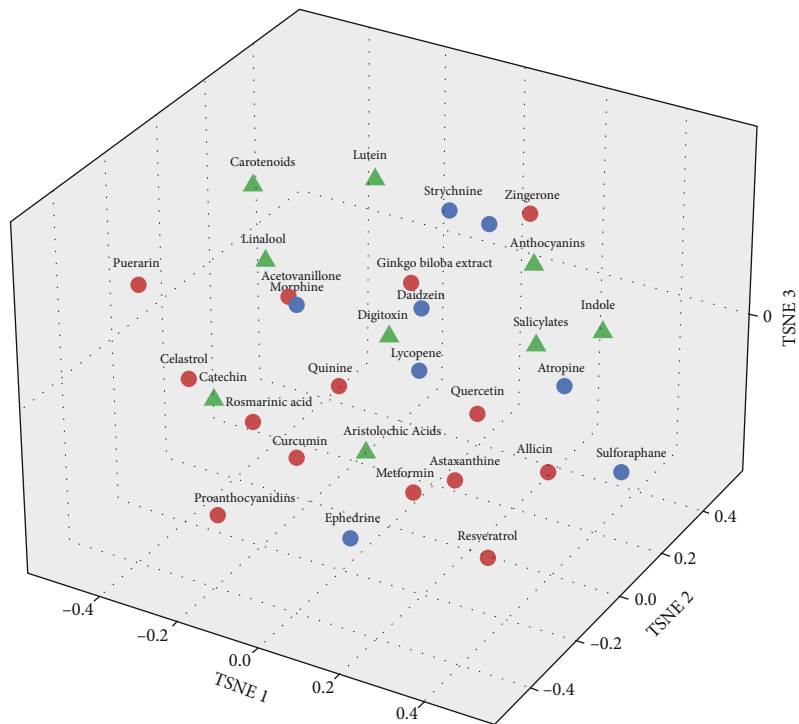
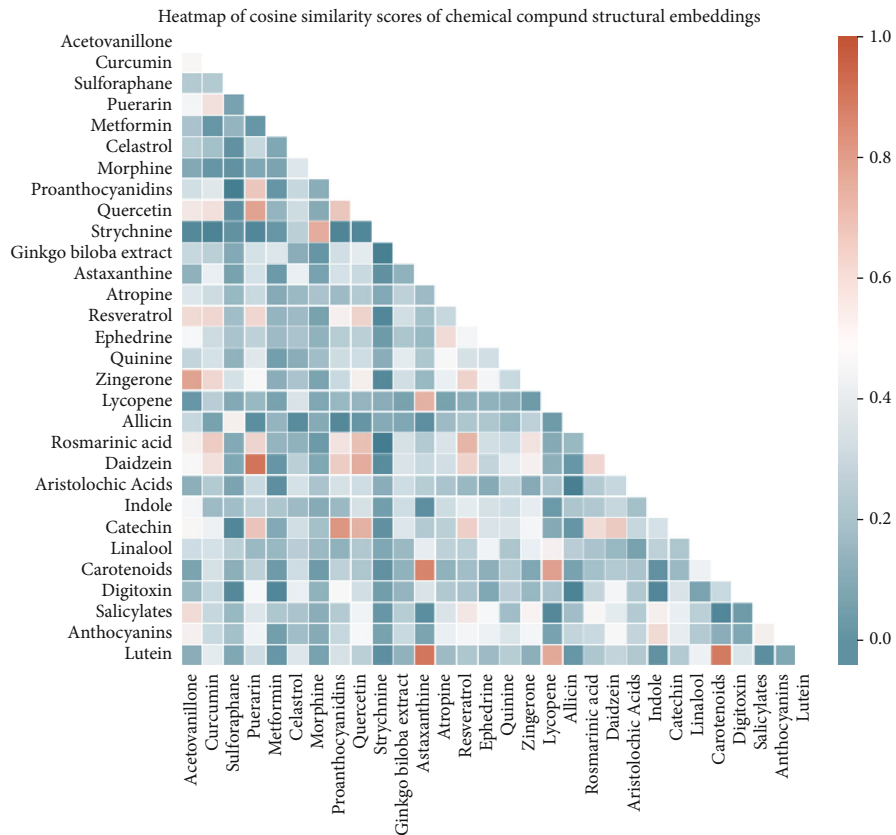
For quantitative examination of VC severity, we used the acid elution method; 0.6 N hydrochloric acid was added to the treated and control group cells and sustained overnight followed by phosphate-buffered saline (PBS) wash. A calcium detection kit (ab102505; Abcam, Cambridge, UK) was then applied, in which chromogenic agents were admixed with eluents. We detected absorbance at 575 nm using a spectrophotometric reader, and results were normalized to cell counts obtained after trypsinization by trypan blue counting using a hemocytometer.

In order to test whether selected chemicals exhibited potential for treating VC, we compared calcification quantitation results between the control group, the calcified group, and the calcified group treated with the designated molecules, using Student’s *t*-test. Sulforaphane, daidzein, and catechin were obtained from Sigma (SI-S6317, SI-D7802, and SI-C1788; St Louis, MO, USA). For calcium quantitation, experiments were obtained in repeated experiments.

3. Results

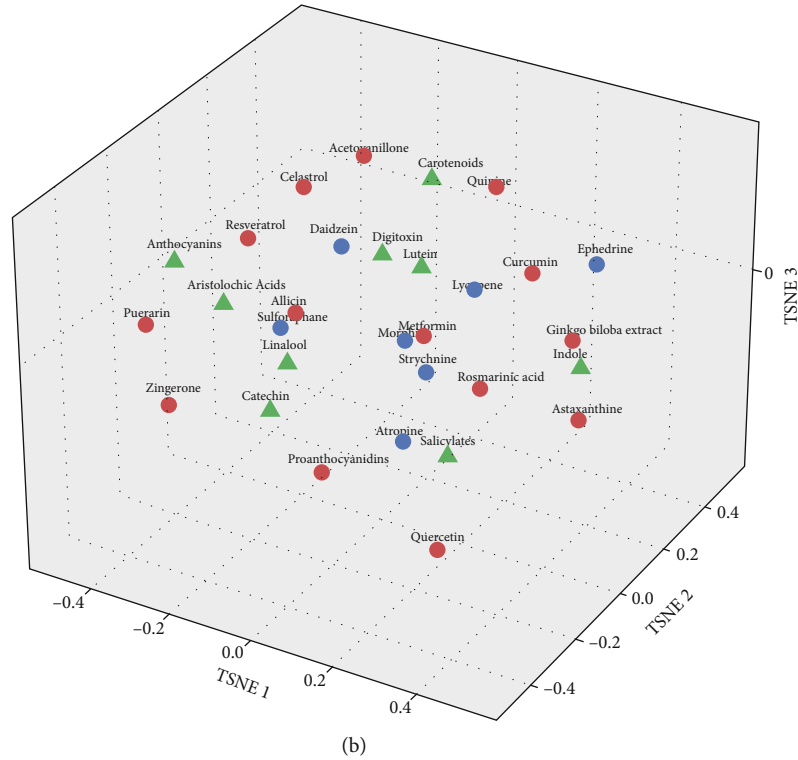
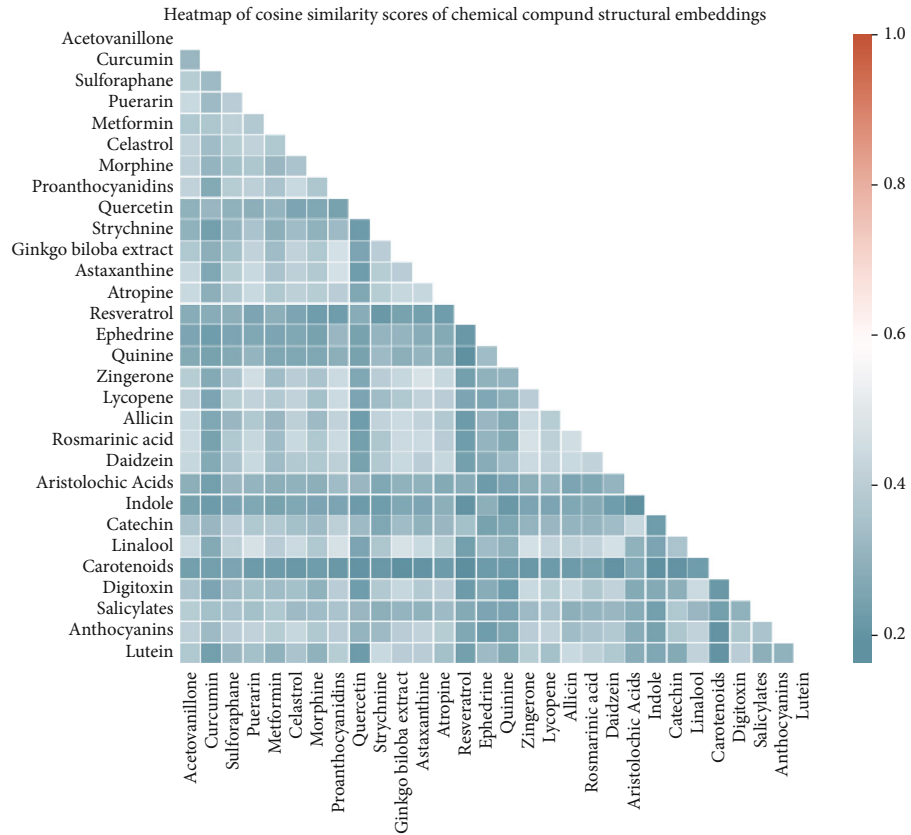
From the cross-linked database, we collected 6,790 drugs with available SMILES data in PubChem and 11,958 GO terms to analyze the features of compounds. A local embedding vector among nodes learned by an end-to-end transformer network and the *node2vec* algorithm and aggregated the information for global embedding vectors from a heterogeneous network via the graph neural network. A binary classification algorithm of drug-disease associations was conducted for triaging compounds. We focused on natural compounds from plants as the sources of new therapeutic drug candidates, a strategy also adopted by others [31]. For test candidate selection, we referred to the CTD database, in which compounds exhibiting a direct evidence for a drug-disease association were marked as having a potentially “therapeutic role.” For other compounds not listed in the CTD database, we considered them to have unknown drug-disease associations, and these compounds were assigned a probability score predicted by our model.

We randomly selected 30 plant-derived compounds after an extensive review of phytomedicinal reports [32], in conjunction with results derived from structure-only embedding vectors and those from functional annotation-only embedding vectors. We examined the similarity scores between any two out of these 30 compounds to see the diversity among different groupings. This was visually displayed by applying t-SNE plots with different feature embedding vectors supported and observing whether compounds belonging to the same category might come close together but different categories remained separated in the t-SNE plots. We found it difficult to differentiate between different categories of compounds in the t-SNE plots based only on chemical structure (SMILES) embedding vectors (Figure 4(a), A) or functional annotations feature embedding vectors (Figure 4(b), A), since the similarity scores were mostly low between any pair of compounds. Some of the plant-derived compounds with cosine similarity of their structural embedding vectors were above 0.85, but similar structural embedding vectors of the drug compounds did not always reflect similar functional bioactivities in general (Figures 4(a), A and 4(b), A). Clusters of potential compounds (red circles, green triangles, and blue circles) tended to mix together without a clear distinction in the 3-dimensional t-SNE plots (Figures 4(a), B and 4(b), B). The first two t-SNE dimensions did not provide a clear separation between clusters, especially when labels were not used for the t-SNE algorithm. It is likely that the biological associations between drugs and diseases are quite complex and that unidimensional groupings based on structure or function-based strategies may not be able to reveal the exact landscape of patterns responsible for the observed drug-disease associations. On the other hand, the cosine similarity scores among our global embedding vectors became more diverse than those using structure or functional vectors (Figure 4(c), A). Ours conferred a greater distinction between potential compounds, presenting as higher similarity scores for the same vector-based categories of compounds but lower scores for different compound categories.



(a)

FIGURE 4: Continued.



(b) FIGURE 4: Continued.

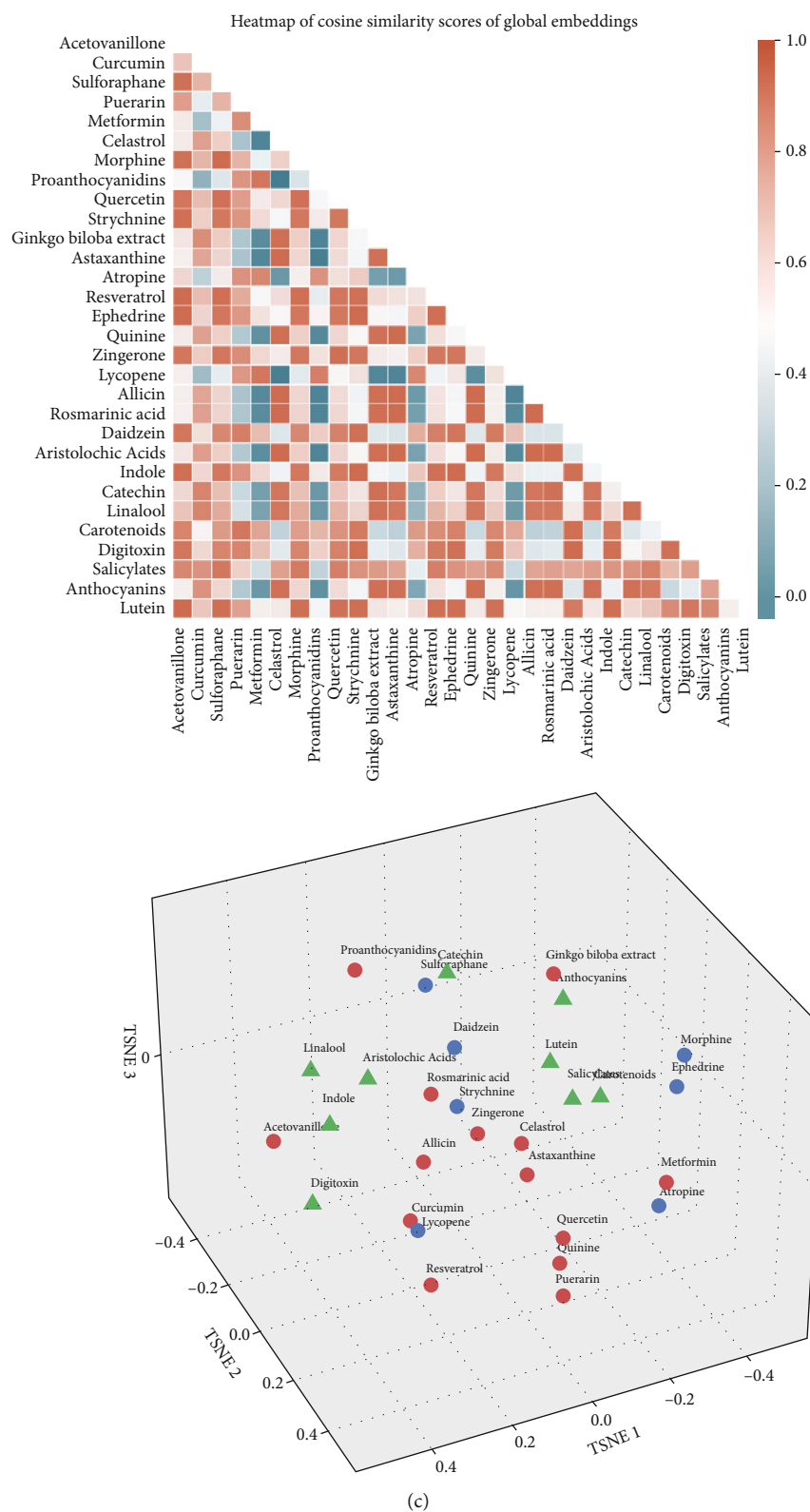
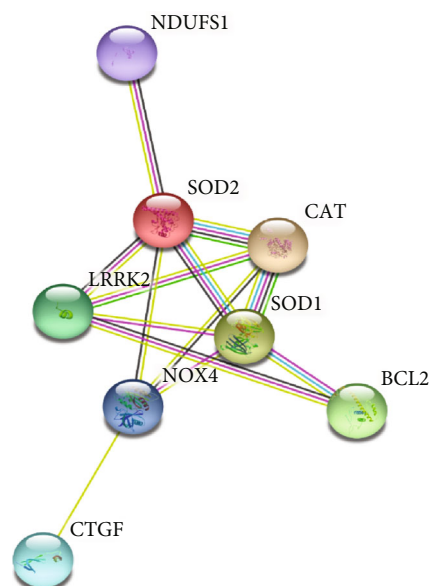


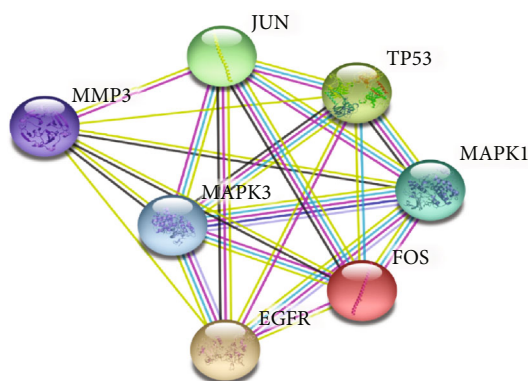
FIGURE 4: The cosine similarity scores and t-SNE projection of different feature representations among 30 plant-derived compounds based on (a) chemical SMILES embedding vectors, (b) GO functional embedding vectors, and (c) global embedding vectors. Compounds with a higher probability score predicted by our model, and also, direct evidences in CTD and published reports were shown in red circles. This with a higher probability score but without evidence supported were in blue circles, while the rest with a low probability score were represented by green triangles. CTD: Comparative Toxicogenomic Database; GO: Gene Ontology; SMILES: Simplified Molecular-Input Line-Entry System; t-SNE, *t*-distributed stochastic neighbor embedding.

DrugID	DrugName	Probability
C056165	Acetovanillone	0.98
D003474	Curcumin	0.88
C016766	Sulforaphane	0.86
C033607	Puerarin	0.78
D008687	Metformin	0.78
C050414	Celastrrol	0.72
D009020	Morphine	0.7
D044945	Proanthocyanidins	0.68
D011794	Quercetin	0.68
D013331	Strychnine	0.68
C063170	Ginkgo Biloba Extract	0.64
C005948	Astaxanthin	0.64
D001285	Atropine	0.64
D000077185	Resveratrol	0.62
D004809	Ephedrine	0.62
D011803	Quinine	0.6
C013738	Zingerone	0.58
D000077276	Lycopene	0.58
C006452	Allicin	0.56
C041376	Rosmarinic acid	0.56
C004742	Daidzein	0.55
D034341	Aristolochic acids	0.46
C030374	Indole	0.44
D002392	Catechin	0.4
C018584	Linalool	0.38
D002338	Carotenoids	0.38
D004074	Digitoxin	0.32
D012459	Salicylates	0.3
D000872	Anthocyanins	0.22
D014975	Lutein	0.12

(a)



(b)



(c)

FIGURE 5: Results of our machine learning-based model including (a) the probability scores of the 30 plant compounds predicted by our model and the interactome analyses of (b) sulforaphane and (c) daidzein regarding relevant expressions in pathogenic molecules involved in VC.

Results on a t-SNE plot illustrated the better differentiation ability between 3 groups of compounds (Figure 4(c), B) compared to the traditional approaches (Figures 4(a), B and 4(b), B).

The probability scores of the 30 plant compounds predicted by our classification model are listed in Figure 5(a). Three categories were recognized, including those with possible drug-disease relationship (probability scores > 0.5) and were affirmed in CTD or existing reports (shown in red),

those with possible drug-disease relationship but never tested before (shown in blue), and those without drug-disease relationship predicted (score < 0.5) while also without prior evidence (shown in green). Compounds within the red category (Figure 5(a)) have already been validated in the existing literature and were compatible with our prediction results, so we did not choose test candidates from this category. To examine the validity of our model, we purposefully selected three compounds, two from the blue

TABLE 1: The performance comparison with the state-of-the-art drug-disease association prediction methods.

Method tested	<i>F1</i>
LRSSL [36]	0.202
SCMFDD-Chem [37]	0.303
SCMFDD-Domain [37]	0.309
SCMFDD-GO [37]	0.314
Fusion+RF [38]	0.580
Ours	0.724

GO: Gene Ontology; LRSSL: Laplacian regularized sparse subspace learning; RF: random forest; SCMFDD: similarity constrained matrix factorization method for the drug-disease association prediction.

category (possible drug-disease relationship but without prior evidence) and one from the green category (with low probability and without evidence, either) for experimental validation (Figure 5(a)). The former group included sulforaphane (probability score 0.86, the highest ranked one in blue category) and daidzein (probability score 0.54, the lowest ranked one in blue category), and the latter group included catechin (probability score 0.4, in green category) (Figure 5(a)). The selection of sulforaphane and daidzein was based on the presumption that those with the highest and lowest probability scores in the blue category could encompass the entire spectrum of candidates with plausible therapeutic efficacy but without experimental evidence.

Sulforaphane is an antioxidative nutraceutical enriched in cruciferous vegetables [33], and daidzein is a plant-origin isoflavone [34]. Catechin, a green tea constituent [35], serves as a comparator with an estimated neutral effect. None of these molecules has been tested for effects against VC in the existing literature. Functional annotation and interactome analyses showed that sulforaphane (Figure 5(b)) and daidzein (Figure 5(c)) could influence the expressions of various genes that played an important pathogenic role during the course of VC [6].

3.1. Comparing Our Approach to Other Machine Learning or Predictive Approaches. We evaluated the performance of our approach by comparing our results with those from the state-of-the-art drug-disease association prediction methods. We adopted the Laplacian regularized sparse subspace learning (LRSSL) benchmark dataset including 3,051 drug-disease associations between 763 FDA-approved drugs and 681 diseases [36]. We constructed three models by combining one of the drug similarities metrics of the chemical substructures (SCMFDD-Chem), protein domains of drug-targeted proteins (SCMFDD-Domain), gene ontology information of drug-targeted proteins (SCMFDD-GO), and disease semantic similarity metrics in SCMFDD method [37]. We adopted *F1*-measure (*F1*) as the primary metric, which was defined as the harmonic mean of recall and precision in highly skewed gold standard benchmark datasets. The results, outlined in Table 1, showed that our approach achieved a *F1* of 0.724, which was higher than those from the previous studies. These data strongly support the reliability of our model in predicting drug-disease interactions.

Previous studies mainly focused on matrix factorization methods for drug or disease similarity matrix. We also applied the fusion approach introduced by Wang et al. [39] to all drug-related and disease-related similarity matrices to construct two fused drug and disease similarity matrices. Then, we used traditional machine learning classification models such as random forest for drug-disease association prediction. *F1*-measure (*F1*) was used as the primary metric as well. Results are outlined in Table 1, which showed that the representation learning using GNN could capture more semantic meaning of drug and disease than similarity metrics fusion method could do.

3.2. Testing the Therapeutic Efficacy of Selected Candidates. Prominent calcification could be observed in VSMCs subjected to conditioned media, both under gross and microscopic examinations (Figures 6(a) and 6(b)). VC *in vitro* manifested as reddish-brown appearing calcification nodules scattered throughout the examined fields (Figures 6(a) and 6(b)). Quantitatively, a 5- to 8-fold increase in the amount of calcium deposited occurred in the calcification group compared to that in the control group (Figures 6(a)–6(c)).

After adding 0.1 μ M sulforaphane, significantly fewer calcification nodules were found under gross and microscopic examination, with around 40% to 50% reduction in the eluted calcium amount (Figure 6(a)). Similarly, 10 μ M daidzein prominently attenuated the extent of calcification in treated VSMCs, with 30% to 40% lower calcium deposition compared to the nontreated group (Figure 6(b)). However, when adding catechin to calcified VSMCs ranging from 5 to 10 μ M, there was no significant change in the extent of calcification nodule development (Figure 6(c)). Calcium quantitation results also revealed insignificant differences between catechin-treated and nontreated groups (Figure 6(c)).

4. Discussion

In the current study, we combined a deep learning approach and experimental studies to unravel potential therapeutic compounds capable for treating VC. We harnessed multiple data sources, including CTD, DrugBank, PubChem, OMIM, and BioGrid, for retrieving chemical structures, genomic/protein/functions/disease relationships and their mutual interactions to expand the biologic landscape being integrated. CTD, as the anchor source for platform integration, suited our aim ideally since CTD is a prestigious literature-based database containing an extensive list of chemicals, genomics, phenomics, diseases, and exposomic information [22]. Through representation learning from deep neural network architectures, we curated a machine learning model for scoring to each compound regarding their probability of having a therapeutic relationship with VC. We subsequently tested candidate compounds with a plausible effect and those with a low probability in a well-established *in vitro* VC model. We were able to show that the predicted compounds worked well, while low probability compound exhibited a neutral effect. Our findings can be inspiring both from the methodological comprehensiveness and the validity of

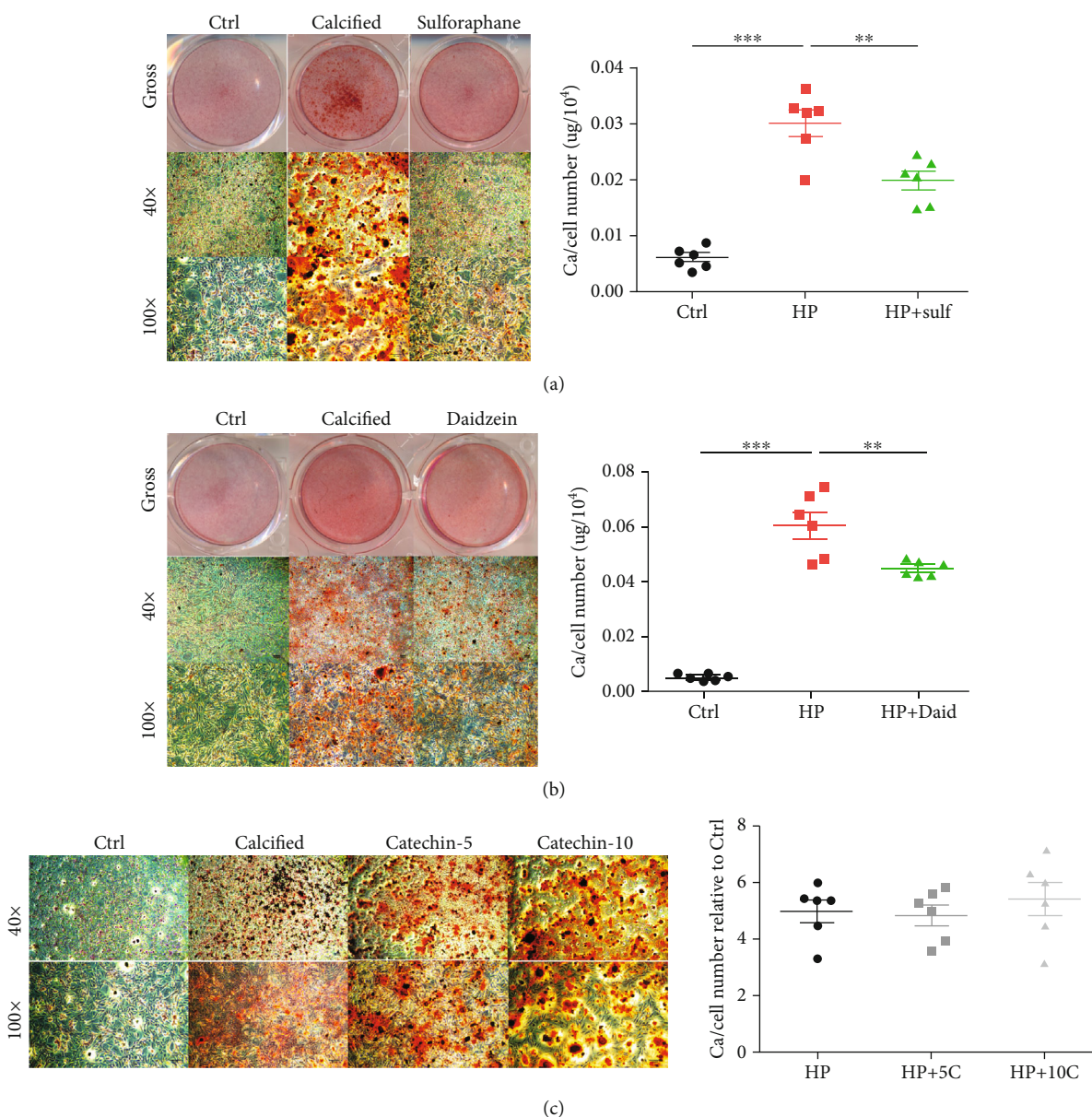


FIGURE 6: Validating the positively predicted candidates (a) sulforaphane, (b) daidzein, and (c) catechin, a candidate with a neutral effect in an *in vitro* vascular calcification model. Calcification nodules were stained using Alizarin Red staining and imaged without (upper row) and with different magnification levels (mid and lower rows) under microscopic examinations. Comparisons of calcium deposition quantities were done between groups using Student's *t*-tests. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. C: catechin; Ctrl: control; Daid: daidzein; HP: high phosphate; Sulf: sulforaphane.

predicted compounds and serve as an example for the utility of deep learning in identifying promising agents to treat vasculopathy. The representation learning and deep learning model can be a quick and inexpensive computational screening tool to assist in the early drug discovery process.

Previous studies integrated multiple sources of drugs, genomic, and disease phenotype features such as drug target proteins, side effects, chemical fingerprints and omics data, and the network measurements of the interaction network for drug-disease interaction prediction. Moreover, existing approaches considered the similarity information of drug or disease for investigating common characteristics without considering the topological information between drugs and

diseases [36, 37, 40–42]. Therefore, there is ample need for developing a graph-based model to capture the heterogeneous information from the network. Therefore, in this study, we established and validated a representation learning model using GNN for drug-disease association prediction in order to achieve drug repurposing.

Effective drug screening and applicability repurposing remains a daunting task for scientists. A “chemogenomic” approach has long been used to predict potential drug compounds that exhibit efficacy against certain diseases/phenotypes based on chemical-protein interactions and structure-activity relationships [43]. Since the inception of high-throughput screening, the popular approach evolves

to studying diseases/phenotypes using mass profiling of cellular/tissue/biological fluid samples from diseased and non-diseased populations, followed by data alignment and comparison, screening for differentially expressed genes/mRNAs/noncoding RNAs/proteins/metabolites, effect size ranking, with or without further input from other adjunct data sources. Results from this newer approach assist greatly in uncovering influential signaling pathways or mediators that count for the disease/phenotype of interest [44]. This strategy can be credited for its straightforwardness and speed in pathogenesis investigation, and its results expectedly facilitate the choice of candidates targeting the newly identified pathways. However, a vital linkage between the identified pathogenic signaling and the selection of candidate drugs is frequently made arbitrarily, attenuating the probability of uncovering novel treatment molecules. Moreover, the conventional high-throughput screening approach rarely addresses the role of compound structures in mechanistic explanations and forgoes the chance of yielding more drug candidates if researchers put little emphasis on the importance of chemical structures in drug screening [45]. In this study, we planned to reap the benefits from both the chemogenomic approach and the phenotype-based high-throughput screening approach, through creating an arching linkage between the structure-activity-relationship and amassing big data from multiple omic disciplines, accomplished by machine learning. A similar approach has been promoted for antineoplastic drug and more recently in antiviral drug virtual screening [46]. In addition, the process of drug repurposing for VC can be further sharpened by combining therapeutic responses with molecular markers for classifying the severity of VC [47, 48].

The strength of our strategy lies in the application of deep learning to selecting therapeutic compounds, an approach that contrasts sharply with other studies harnessing machine learning and big data in the field of VC. Indeed, a recent state-of-the-art review about machine learning and artificial intelligence applicability in cardiovascular calcification addresses mostly calcification imaging and quantification, early diagnosis, and outcome prediction [49]. Regarding new therapeutic candidate screening, researchers rarely utilized the machine learning-based approach. Luechtefeld and colleagues used an expanded database of candidate molecules in combination with toxicological features extracted from sources of regulatory authorities to predict whether candidates exhibit health hazards [50]. They derived fair results based on such approach. However, VC-targeting therapeutic compound screening using a network medicine and machine learning-based approach had not been attempted before. From this perspective, we believe that our findings have their merits and shed light on future VC treatments.

The efficacy of three compounds predicted by our algorithm was validated in this study. Sulforaphane is a metabolite originating from the myrosinase-assisted hydrolysis of glucoraphanin, which is predominantly found in broccoli cultivars and rapidly absorbed into the body [51]. Sulforaphane has been shown to arrest cell cycles through modulating cyclin expressions, inducing cancer cell apoptosis through

altering redox balance and activating caspases, inhibiting histone deacetylase, and inducing autophagy [52]. Sulforaphane also activates nuclear factor-erythroid 2-related factor 2 (Nrf2) and indirectly suppresses nuclear factor- κ B (NF- κ B) activity, exhibiting a salutary effect in CKD, an important VC risk factor [53]. Prior studies showed that sulforaphane reduced VSMC oxidative stress and attenuated VSMC proliferation as well as phenotype switching [54]. Daidzein, a soybean-derived isoflavone and phytoestrogen, has been known to possess estrogenic effect through its bacterial metabolite equol, anti-inflammatory effect, and antiosteoporotic efficacy [55]. Daidzein, along with other phytoestrogens like genistein, was shown to ameliorate DNA damages in VSMCs and decrease neointima proliferation [56, 57], but none evaluated its effect on VSMC calcification. We were able to demonstrate for the first time that sulforaphane and daidzein could ameliorate VC in addition to VSMC proliferation/migration inhibition, which has not been reported before. These findings support the utility of our cross-platform algorithm. On the other hand, catechins, the main antioxidants contained in green tea, exhibit variable degrees of reactive oxygen species neutralization and are touted to have benefits against cancer and neurodegenerative disorders [58]. Our algorithm successfully identified this class of compounds as having a lower probability of working against VC, supporting the differentiation ability of our algorithm.

According to experiences from synthetic biology, especially structure-based predictions based on machine learning techniques similar to ours, synthetic bioactive compounds may not necessarily exhibit the same biologic effects (therapeutic or toxicity ones) as their parent compounds [59]. Moreover, the components of incorporated database and network references in the prediction models also influence the applicability of findings across different compounds. Nonetheless, we have integrated a wide array of biomedical database, containing natural neutraceuticals/pharmaceuticals and synthetic bioactive compounds as well. We believe that results from our machine learning-based prediction model can be applicable to both natural and synthetic compounds. In addition, among the three natural compounds tested in this study, sulforaphane has been shown to exhibit a fair oral bioavailability, while daidzein, a flavonoid with polyphenol structure and catechin, both have poor oral bioavailability and possibly suboptimal absorption [60–62]. Theoretically, the oral administration of these compounds may suffer from inadequate therapeutic efficacy partially attributable to these pharmacokinetic disadvantages. However, newer techniques have emerged to facilitate the absorption and distribution of these bioactive compounds, such as liposomal formulation and other nanocarrier-loaded formulation. Therefore, the therapeutic potential of compounds identified in this study remains promising in the future.

Our study has its strengths and limitations. The integration of multiple database and dimensions for virtual drug screening greatly enhanced our ability to detect potentially useful therapeutic compounds for VC. The verification of such effects in experimental models supported the credibility of our algorithm. However, several remaining issues needed

further adjudication. First, we did not verify the therapeutic responses of these compounds using *in vivo* models, and there is a possibility that these compounds may not survive *in vivo* tests. Second, VC is a heterogeneous disease entity, and it is plausible that an individual compound may work against VC of a certain pathogenic subtype only but not others. Finally, our algorithm was not expected to be fixed, since the inclusion of more dimensions in data integration or other data sources might sharpen the prediction ability and derive different sets of prediction. More are still needed in order to pursue clinically meaningful treatment for VC.

5. Conclusion

In conclusion, we established a prediction algorithm for identifying novel therapeutic compounds against VC, by integrating megadatabase including CTD, DrugBank, PubChem, OMIM, and BioGrid, incorporating chemogenomic components and multiple omic data sources. Several potential candidates were predicted as being efficacious based on scoring thresholds. We validated three out of the predicted compounds, 2 with a positive effect and 1 with a neutral effect, and affirmed our prediction results in experimental studies. We believe that our approach can be credible and facilitate the subsequent selection of optimal therapeutic compounds for VC, a dreadful vascular complication with gender differences and complex pathophysiology [63, 64].

Abbreviations

AUC:	Area under the receiver operating characteristics curve
CKD:	Chronic kidney disease
CTD:	Comparative Toxicogenomic Database
DM:	Diabetes mellitus
DMEM:	Dulbecco's modified Eagle's medium
FBS:	Fetal bovine sera
FGF-23:	Fibroblast growth factor-23
GNN:	Graph deep neural network architectures
GO:	Gene Ontology
IS:	Indoxyl sulfate
NF- κ B:	Nuclear factor- κ B
Nrf2:	Nuclear factor-erythroid 2-related factor 2
PBS:	Phosphate-buffered saline
SMILES:	Simplified Molecular-Input Line-Entry System
t-SNE:	<i>T</i> -distributed stochastic neighbor embedding
VC:	Vascular calcification
VSMC:	Vascular smooth muscle cell.

Data Availability

The raw data for conducting this analysis are available upon reasonable request to the corresponding author.

Consent

Consent was not necessary.

Disclosure

The sponsors have no role in the study design, data collection, analysis, and result interpretation of this study.

Conflicts of Interest

The authors have no relevant financial or nonfinancial competing interests to declare in relation to this manuscript.

Authors' Contributions

Study design was done by CTC and HYY. Data analysis was conducted by CTC, YTT, WTL, and HYY. Article drafting was performed by CTC, YTT, WTL, HYY, and CKC. All authors approved the final version of the manuscript.

Acknowledgments

We are grateful to the Second Core Laboratory, Department of Medical Research of National Taiwan University Hospital and the National Taiwan University Center of Genomic and Precision Medicine for their technical input. The study is financially sponsored by the National Taiwan University Hospital BeiHu Branch (11001) and Ministry of Science and Technology, Taiwan (MOST 109-2314-B-002-193-MY3).

References

- [1] S. Zununi Vahed, S. Mostafavi, S. M. Hosseiniyan Khatibi, M. M. Shoja, and M. Ardalani, "Vascular calcification: an important understanding in nephrology," *Vascular Health and Risk Management*, vol. Volume 16, pp. 167–180, 2020.
- [2] M. Jadidi, W. Poulson, P. Aylward et al., "Calcification prevalence in different vascular zones and its association with demographics, risk factors, and morphometry," *American Journal of Physiology Heart and Circulatory Physiology*, vol. 320, no. 6, pp. H2313–h2323, 2021.
- [3] Z. Yao, C. Wang, Q. Zhang, S. Ma, B. Gui, and C. Duan, "Prevalence of abdominal artery calcification in dialysis patients with end-stage renal disease: a systematic review and meta-analysis," *International Urology and Nephrology*, vol. 49, no. 11, pp. 2061–2069, 2017.
- [4] X. R. Wang, J. J. Zhang, X. X. Xu, and Y. G. Wu, "Prevalence of coronary artery calcification and its association with mortality, cardiovascular events in patients with chronic kidney disease: a systematic review and meta-analysis," *Renal Failure*, vol. 41, no. 1, pp. 244–256, 2019.
- [5] C. T. Chao, S. Y. Lee, J. Wang, K. L. Chien, and K. Y. Hung, "The risk trajectory of different cardiovascular morbidities associated with chronic kidney disease among patients with newly diagnosed diabetes mellitus: a propensity score-matched cohort analysis," *Cardiovascular Diabetology*, vol. 20, no. 1, p. 86, 2021.
- [6] C. T. Chao, H. Y. Yeh, Y. T. Tsai et al., "Natural and non-natural antioxidative compounds: potential candidates for treatment of vascular calcification," *Cell Death Discovery*, vol. 5, no. 1, p. 145, 2019.
- [7] C. T. Chao and S. H. Lin, "Uremic vascular calcification: the pathogenic roles and gastrointestinal decontamination of uremic toxins," *Toxins*, vol. 12, no. 12, p. 812, 2020.

- [8] Y. C. Hou, C. L. Lu, T. H. Yuan, M. T. Liao, C. T. Chao, and K. C. Lu, "The epigenetic landscape of vascular calcification: an integrative perspective," *International Journal of Molecular Sciences*, vol. 21, no. 3, p. 980, 2020.
- [9] Y. Chen, X. Zhao, and H. Wu, "Arterial stiffness: a focus on vascular calcification and its link to bone mineralization," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 40, no. 5, pp. 1078–1093, 2020.
- [10] C. T. Chao, H. Y. Yeh, Y. T. Tsai, C. K. Chiang, and H. W. Chen, "A combined microRNA and target protein-based panel for predicting the probability and severity of uraemic vascular calcification: a translational study," *Cardiovascular Research*, vol. 117, no. 8, pp. 1958–1973, 2021.
- [11] C. T. Chao, Y. P. Liu, S. F. Su et al., "Circulating microRNA-125b predicts the presence and progression of uremic vascular calcification," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 37, no. 7, pp. 1402–1414, 2017.
- [12] T. Isakova, T. L. Nickolas, M. Denburg et al., "KDOQI US commentary on the 2017 KDIGO clinical practice guideline update for the diagnosis, evaluation, prevention, and treatment of chronic kidney disease-mineral and bone disorder (CKD-MBD)," *American Journal of Kidney Diseases*, vol. 70, no. 6, pp. 737–751, 2017.
- [13] C. Wanner, "Moderator's view: treatment of vascular calcification is a physical impossibility, so far," *Nephrology, Dialysis, Transplantation*, vol. 30, no. 3, pp. 358–359, 2015.
- [14] J. P. Schanstra, T. T. Luong, M. Makridakis et al., "Systems biology identifies cytosolic PLA2 as a target in vascular calcification treatment," *JCI Insight*, vol. 4, no. 10, article e125638, 2019.
- [15] C. T. Chao, H. Y. Yeh, Y. T. Tsai et al., "Astaxanthin counteracts vascular calcification in vitro through an early up-regulation of SOD2 based on a transcriptomic approach," *International Journal of Molecular Sciences*, vol. 21, no. 22, p. 8530, 2020.
- [16] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.
- [17] T. H. Rim, C. J. Lee, Y. C. Tham et al., "Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs," *Lancet Digit Health*, vol. 3, no. 5, pp. e306–e316, 2021.
- [18] J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug-disease relationships for computational drug repositioning," *Briefings in Bioinformatics*, vol. 12, no. 4, pp. 303–311, 2011.
- [19] J. T. Dudley, E. Schadt, M. Sirota, A. J. Butte, and E. Ashley, "Drug discovery in a multidimensional world: systems, patterns, and networks," *Journal of Cardiovascular Translational Research*, vol. 3, no. 5, pp. 438–447, 2010.
- [20] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang, "Heterogeneous network embedding via deep architectures," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia, 2015.
- [21] M. Iskar, G. Zeller, X. M. Zhao, V. van Noort, and P. Bork, "Drug discovery in the age of systems biology: the rise of computational approaches for data integration," *Current Opinion in Biotechnology*, vol. 23, no. 4, pp. 609–616, 2012.
- [22] A. P. Davis, C. J. Grondin, R. J. Johnson et al., "The Comparative Toxicogenomics Database: update 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D948–D954, 2019.
- [23] V. Law, C. Knox, Y. Djoumbou et al., "DrugBank 4.0: shedding new light on drug metabolism," *Nucleic Acids Research*, vol. 42, no. D1, pp. D1091–D1097, 2014.
- [24] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," *Journal of Computational Chemistry*, vol. 38, no. 16, pp. 1291–1307, 2017.
- [25] P. Morris, R. St Clair, W. E. Hahn, and E. Barenholtz, "Predicting binding from screening assays with transformer network embeddings," *Journal of Chemical Information and Modeling*, vol. 60, no. 9, pp. 4191–4199, 2020.
- [26] A. Grover and J. Leskovec, "node2vec: scalable feature learning for networks," *Kdd*, vol. 2016, pp. 855–864, 2016.
- [27] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019.
- [28] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 2017.
- [29] Z. Gong, P. Zhong, and W. Hu, "Diversity in machine learning," *IEEE Access*, vol. 7, pp. 64323–64350, 2019.
- [30] G. Maggiora, M. Vogt, D. Stumpfe, and J. Bajorath, "Molecular similarity in medicinal chemistry," *Journal of Medicinal Chemistry*, vol. 57, no. 8, pp. 3186–3204, 2014.
- [31] D. M. Muntean, A. Sturza, I. Z. Pavel, and O. M. Duicu, "Modulation of cancer metabolism by phytochemicals - a brief overview," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 18, no. 5, pp. 684–692, 2018.
- [32] A. G. Atanasov, B. Waltenberger, E. M. Pferschy-Wenzig et al., "Discovery and resupply of pharmacologically active plant-derived natural products: a review," *Biotechnology Advances*, vol. 33, no. 8, pp. 1582–1614, 2015.
- [33] R. T. Ruhee and K. Suzuki, "The integrative role of sulforaphane in preventing inflammation, oxidative stress and fatigue: a review of a potential protective phytochemical," *Antioxidants*, vol. 9, no. 6, p. 521, 2020.
- [34] D. Das, S. Sarkar, J. Bordoloi, S. B. Wann, J. Kalita, and P. Manna, "Daidzein, its effects on impaired glucose and lipid metabolism and vascular inflammation associated with type 2 diabetes," *BioFactors*, vol. 44, no. 5, pp. 407–417, 2018.
- [35] A. N. Dos Santos, T. R. de L Nascimento, B. L. Gondim et al., "Catechins as model bioactive compounds for biomedical applications," *Current Pharmaceutical Design*, vol. 26, no. 33, pp. 4032–4047, 2020.
- [36] X. Liang, P. Zhang, L. Yan et al., "LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning," *Bioinformatics*, vol. 33, no. 8, pp. 1187–1196, 2017.
- [37] W. Zhang, X. Yue, W. Lin et al., "Predicting drug-disease associations by using similarity constrained matrix factorization," *BMC Bioinformatics*, vol. 19, no. 1, p. 233, 2018.
- [38] R. Hoehndorf, A. Oellrich, D. Rebholz-Schuhmann, P. N. Schofield, and G. V. Gkoutos, "Linking PharmGKB to phenotype studies and animal models of disease for drug repurposing," *Pacific Symposium on Biocomputing*, vol. 2012, pp. 388–399, 2012.
- [39] B. Wang, A. M. Mezlini, F. Demir et al., "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, no. 3, pp. 333–337, 2014.
- [40] A. Gottlieb, G. Y. Stein, E. Ruppim, and R. Sharan, "PREDICT: a method for inferring novel drug indications with application to personalized medicine," *Molecular Systems Biology*, vol. 7, no. 1, p. 496, 2011.

- [41] Y. Wang, S. Chen, N. Deng, and Y. Wang, "Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data," *PLoS One*, vol. 8, no. 11, article e78518, 2013.
- [42] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinformatics*, vol. 34, no. 11, pp. 1904–1912, 2018.
- [43] L. Jacob and J. P. Vert, "Protein-ligand interaction prediction: an improved chemogenomics approach," *Bioinformatics*, vol. 24, no. 19, pp. 2149–2156, 2008.
- [44] Y. Hizukuri, R. Sawada, and Y. Yamanishi, "Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner," *BMC Medical Genomics*, vol. 8, no. 1, p. 82, 2015.
- [45] Q. Li and S. Shah, "Structure-based virtual screening," *Methods in Molecular Biology*, vol. 1558, pp. 111–124, 2017.
- [46] R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar, "Artificial intelligence to deep learning: machine intelligence approach for drug discovery," *Molecular Diversity*, vol. 25, no. 3, pp. 1315–1360, 2021.
- [47] C. T. Chao, H. Y. Yeh, D. S. Han, J. W. Huang, and K. C. Huang, "Determinants of circulating microRNA-125b, a risk predictor of vascular calcification, among community-dwelling older adults<sup/>," *Clinical and Translational Medicine*, vol. 10, no. 4, article e145, 2020.
- [48] C. T. Chao, D. S. Han, and J. W. Huang, "Circulating microRNA-125b levels are associated with the risk of vascular calcification in healthy community-dwelling older adults," *Front Cardiovasc Med*, vol. 8, article 624313, 2021.
- [49] M. A. Rogers and E. Aikawa, "Cardiovascular calcification: artificial intelligence and big data accelerate mechanistic discovery," *Nature Reviews. Cardiology*, vol. 16, no. 5, pp. 261–274, 2019.
- [50] T. Luechtefeld, D. Marsh, C. Rowlands, and T. Hartung, "Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility," *Toxicological Sciences*, vol. 165, no. 1, pp. 198–212, 2018.
- [51] L. G. West, K. A. Meyer, B. A. Balch, F. J. Rossi, M. R. Schultz, and G. W. Haas, "Glucoraphanin and 4-hydroxyglucobrassicin contents in seeds of 59 cultivars of broccoli, raab, kohlrabi, radish, cauliflower, brussels sprouts, kale, and cabbage," *Journal of Agricultural and Food Chemistry*, vol. 52, no. 4, pp. 916–926, 2004.
- [52] O. A. Elkashty and S. D. Tran, "Sulforaphane as a promising natural molecule for cancer prevention and treatment," *Current Medical Science*, vol. 41, no. 2, pp. 250–269, 2021.
- [53] L. F. Cardozo, L. A. Alvarenga, M. Ribeiro et al., "Cruciferous vegetables: rationale for exploring potential salutary effects of sulforaphane-rich foods in patients with chronic kidney disease," *Nutrition Reviews*, vol. nuaa129, 2021.
- [54] X. He, J. Deng, X.-J. Yu, S. Yang, Y. Yang, and W.-J. Zang, "Activation of M3AChR (type 3 muscarinic acetylcholine receptor) and Nrf2 (nuclear factor erythroid 2-related factor 2) signaling by choline alleviates vascular smooth muscle cell phenotypic switching and vascular remodeling," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 40, no. 11, pp. 2649–2664, 2020.
- [55] M. Uehara, "Isoflavone metabolism and bone-sparing effects of daidzein-metabolites," *Journal of Clinical Biochemistry and Nutrition*, vol. 52, no. 3, pp. 193–201, 2013.
- [56] K. Mizutani, K. Ikeda, T. Nishikata, and Y. Yamori, "Phytoestrogens attenuate oxidative DNA damage in vascular smooth muscle cells from stroke-prone spontaneously hypertensive rats," *Journal of Hypertension*, vol. 18, no. 12, pp. 1833–1840, 2000.
- [57] S. Liu, L. Wang, W. Wang et al., "TSC-36/FRP inhibits vascular smooth muscle cell proliferation and migration," *Experimental and Molecular Pathology*, vol. 80, no. 2, pp. 132–140, 2006.
- [58] C. Musial, A. Kuban-Jankowska, and M. Gorska-Ponikowska, "Beneficial properties of green tea catechins," *International Journal of Molecular Sciences*, vol. 21, no. 5, p. 1744, 2020.
- [59] L. Pu, M. Naderi, T. Liu, H. C. Wu, S. Mukhopadhyay, and M. Brylinski, "eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates," *BMC Pharmacology and Toxicology*, vol. 20, no. 1, p. 2, 2019.
- [60] A. Tavakkoli, M. Iranshahi, S. H. Hasheminezhad, A. W. Hayes, and G. Karimi, "The neuroprotective activities of natural products through the Nrf2 upregulation," *Phytotherapy Research*, vol. 33, no. 9, pp. 2256–2273, 2019.
- [61] S. V. Luca, I. Macovei, A. Bujor et al., "Bioactivity of dietary polyphenols: the role of metabolites," *Critical Reviews in Food Science and Nutrition*, vol. 60, no. 4, pp. 626–659, 2020.
- [62] H. M. Ezzat, Y. S. R. Elnaggar, and O. Y. Abdallah, "Improved oral bioavailability of the anticancer drug catechin using chitosomes: design, in-vitro appraisal and in-vivo studies," *International Journal of Pharmaceutics*, vol. 565, pp. 488–498, 2019.
- [63] Y. T. Tsai, H. Y. Yeh, C. T. Chao, and C. K. Chiang, "Superoxide dismutase 2 (SOD2) in vascular calcification: a focus on vascular smooth muscle cells, calcification pathogenesis, and therapeutic strategies," *Oxidative Medicine and Cellular Longevity*, vol. 2021, Article ID 6675548, 9 pages, 2021.
- [64] P. Y. Wu, S. Y. Lee, K. V. Chang, C. T. Chao, and J. W. Huang, "Gender-related differences in chronic kidney disease-associated vascular calcification risk and potential risk mediators: a scoping review," *Healthcare*, vol. 9, no. 8, p. 979, 2021.