BioGecko

Vol 12 Issue 03 2023
ISSN NO: 2230-5807

Method for Identifying Herbal Compounds with Anti-Cancer Properties Using Machine Learning Techniques

[*1]Dr. Hanish Singh Jayasingh Chellammal, [2]Dr Shaista Omer, [3]Dr. Deepa Gupta, [3]Dr. Alok Khunteta, 3Dr. Surendra Kumar Swarnkar, [3]Puneet Gupta, [4]Fauzia Bano faruqi

[1]Department of Pharmacology & Pharmaceutical Chemistry, Faculty of Pharmacy, UiTM, Puncak Alam, Selangor-42300, Malaysia.

[2]Principal, TMAE' Society MMJG College of Pharmacy Haveri, Karnataka- 581110.

[3]LBS College of Pharmacy, Tilak Nagar, Jaipur

[4]Department of Applied Chemistry, Aligarh Muslim University, Aligarh, Uttar Pradesh.

*Corresponding Author

*1Dr. Hanish Singh Jayasingh Chellammal

1Department of Pharmacology & Pharmaceutical Chemistry, Faculty of Pharmacy, UiTM, Puncak Alam, Selangor-42300, Malaysia.

hanishsinghjc@gmail.com

Abstract

The present invention discloses a method for identifying herbal compounds with anti-cancer properties using machine learning techniques. The method involves collecting a comprehensive dataset comprising information on herbal compounds, their chemical properties, and experimental data on their anti-cancer activities. Relevant features are extracted from the dataset, including molecular descriptors, physicochemical properties, and structural characteristics of the compounds. The dataset is preprocessed to ensure data quality, and machine learning models, such as support vector machines or neural networks, are trained on the preprocessed data to learn the relationship between the features and the anti-cancer activities.

Introduction

The present invention relates to the field of pharmaceutical research and, in particular, to a method foridentifying herbal compounds with anti-cancer properties using machine learning techniques. More specifically, the invention pertains to a novel approach that utilizes machine learning algorithms to screen and identify herbal compounds that exhibit potential anti-cancer activity[1].

In spite of this extraordinary growth in the cancer biology field, progress in the area of drug discovery remains stagnant, still plagued with long development time to market and exorbitantly high costs despite the systematic implementation of high-throughput screening technologies. To date, bringing a drug to market takes about a decade with research and development (R&D) costs reaching approximately US $2.8 billion (1). Candidate drugs may fail at many points along the drug development pipeline due to numerous reasons such as poor pharmacokinetics, toxicity or lack of clinical efficacy.

A promising solution to the considerable drug development challenges of novel compounds is the use of existing drugs for the treatment of new diseases. Approved drugs have undergone all phases of clinical trials in order to reach the market and thus have a known and accepted safety profile. If a new clinical indication for an approved drug is suggested, that drug may re-enter the clinical trial process at Phase II thus substantially reducing the R&D risk, time and cost (2).

In the past few years, drug repurposing research has benefited greatly by the systematic adoption of computational strategies spanning a large area of unique methodologies and approaches. Molecular modeling of therapeutic protein

targets allows for the understanding of the structural biology as well as high throughput "virtual screenings" to identify novel drug candidates. Advances in machine learning and artificial intelligence (in particular deep learning) provide new insights into how drugs bind to targets and how their physicochemical properties relate to phenotypic changes. In addition, the utilization of such methods also help identify novel anti-cancer targets from the large-scale cancer datasets already collected by multiple initiatives. As the amount of chemical and bioactivity data grows due to the adoption of high-throughput and multi-omics drug profiling assays, these methods are poised to make substantial contributions in cancer drug discovery. Furthermore, the increased accessibility of these public dataset collections further facilitates the potential of computational approaches. These approaches need not be limited in using experimental and biological datasets. The utilization of clinical datasets such as EHR (electronic health records) has also shown promising results. In this review, we discuss the current computational strategies used for drug repurposing in oncology with an emphasis on machine and deep learning.

Background of Research

Cancer is a leading cause of mortality worldwide, necessitating the development of effective treatments. Traditional medicine, including the use of herbal compounds, has been explored for its potential in combating cancer. However, the identification of herbal compounds with anti-cancer properties remains a challenging and time-consuming process due to the vast number of natural products available and the complex interactions within biological systems[2].

Cancer is a complex and devastating disease that continues to be a major global health concern. Despite advancements in treatment modalities, the development of effective anti-cancer therapies remains a significant challenge. In recent years, there has been a growing interest in exploring natural products, including herbal compounds, as potential sources of anti-cancer agents[3].

Traditional medicine systems, such as Ayurveda, Traditional Chinese Medicine (TCM), and traditional healing practices from various cultures, have long relied on herbal remedies for treating a wide range of ailments, including cancer. Herbal compounds derived from medicinal plants often exhibit diverse chemical structures and bioactive properties, making them attractive candidates for drug discovery. However, the identification of herbal compounds with anti-cancer properties is a complex and time- consuming process. Traditional screening methods involve extensive experimentation, which isresource-intensive and often lacks efficiency. Additionally, the vast number of natural products available, along with the intricate interactions within biological systems, further complicates the identification of promising anti-cancer compounds. Conventional screening methods involve extensive experimentation, requiring substantial resources and time. These approaches often lack efficiency and fail to consider the synergistic effects between compounds and their targets. There is a need for an improved method that can effectively and efficiently identify herbal compounds with anti-cancer properties, thereby accelerating the discoveryof potential therapeutic agents[4].

To address these challenges, there is a need for an improved method that can expedite the identification of herbal compounds with anti-cancer properties while considering their complex interactions with cancer cells and their targets. Machine learning techniques offer a promising approach to tackle this problem by harnessing the power of data-driven analysis and predictive modeling[5].

Machine learning algorithms can learn patterns and relationships from large datasets, enabling the discovery of hidden associations between chemical features and anti-cancer activities. By utilizing[6] these algorithms, it becomes possible to screen and prioritize herbal compounds based on their potential efficacy as anti-cancer agents. The integration of machine learning techniques with herbal compound screening provides several advantages. First, it allows for the simultaneous evaluation of a large number of compounds,significantly reducing the time and resources required for screening. Second, machine learning models can capture complex interactions and

synergistic effects between compounds and their targets,leading to more accurate predictions of anti-cancer activity. Third, machine learning-based approaches offer the potential for comprehensive screening by considering multiple molecularfeatures and diverse chemical properties[7].

Materials and methods

The present invention provides a method for identifying herbal compounds with anti-cancer propertiesusing machine learning techniques. The method leverages the power of data-driven analysis and predictive modeling to expedite the screening process and increase the accuracy of compound selection. The present invention aims to address the limitations of conventional screening methods and provide a novel method for identifying herbal compounds with anti-cancer properties using machine learning techniques. This method combines the power of data analysis, feature extraction, model training, and validation to enable efficient and accurate screening of herbal compounds, ultimately accelerating the discovery of potential therapeutic agents for cancer treatment[8].

To further clarify advantages and features of the present disclosure, a more particular description of the invention will be rendered by reference to specific embodiments thereof, which is illustrated in the appended drawings. It is appreciated that these drawings depict only typical embodiments of the invention and are therefore not to be considered limiting of its scope. The invention will be describedand explained with additional specificity and detail with the accompanying drawings[9].

Brief Description of Drawings

These and other features, aspects, and advantages of the present disclosure will become better understood when the following detailed description is read with reference to the accompanying drawings in which like characters represent like parts throughout the drawings, wherein:

Figure 1: illustrates Flowchart illustrating the steps involved in the method for identifying herbal compounds with anti-cancer properties using machine learning techniques[10].

Further, skilled artisans will appreciate that elements in the drawings are illustrated for simplicity andmay not have been necessarily been drawn to scale. For example, the flow charts illustrate the methodin terms of the most prominent steps involved to help to improve understanding of aspects of the present disclosure. Furthermore, in terms of the construction of the device, one or more components of the device may have been represented in the drawings by conventional symbols, and the drawingsmay show only those specific details that are pertinent to understanding the embodiments of the present disclosure so as not to obscure the drawings with details that will be readily apparent to thoseof ordinary skill in the art having benefit of the description herein[11].

Material and methods

For the purpose of promoting an understanding of the principles of the invention, reference will now be made to the embodiment illustrated in the drawings and specific language will be used to describethe same. It will nevertheless be understood that no limitation of the scope of the invention is therebyintended, such alterations and further modifications in the illustrated system, and such further applications of the principles of the invention as illustrated therein being contemplated as would normally occur to one skilled in the art to which the invention relates[12].

It will be understood by those skilled in the art that the foregoing general description and the following detailed description are exemplary and explanatory of the invention and are not intended tobe restrictive thereof.

Reference throughout this specification to “an aspect”, “another aspect” or similar language means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least

one embodiment of the present disclosure. Thus, appearances of the phrase “in an embodiment”, “in another embodiment” and similar language throughout this specification may, but do not necessarily, all refer to the same embodiment[13].

The terms "comprises", "comprising", or any other variations thereof, are intended to cover a non- exclusive inclusion, such that a process or method that comprises a list of steps does not include onlythose steps but may include other steps not expressly listed or inherent to such process or method. Similarly, one or more devices or sub-systems or elements or structures or components proceeded by "comprises...a" does not, without more constraints, preclude the existence of other devices or other sub-systems or other elements or other structures or other components or additional devices or additional sub-systems or additional elements or additional structures or additional components.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The system,methods, and examples provided herein are illustrative only and not intended to be limiting.

Embodiments of the present disclosure will be described below in detail with reference to the accompanying drawings.

The functional units described in this specification have been labeled as devices. A device may be implemented in programmable hardware devices such as processors, digital signal processors, centralprocessing units, field programmable gate arrays, programmable array logic, programmable logic devices, cloud processing systems, or the like. The devices may also be implemented in software forexecution by various types of processors. An identified device may include executable code and may,

for instance, comprise one or more physical or logical blocks of computer instructions, which may, for instance, be organized as an object, procedure, function, or other construct. Nevertheless, the executable of an identified device need not be physically located together, but may comprise disparateinstructions stored in different locations which, when joined logically together, comprise the device and achieve the stated purpose of the device.

Indeed, an executable code of a device or module could be a single instruction, or many instructions, and may even be distributed over several different code segments, among different applications, and across several memory devices. Similarly, operational data may be identified and illustrated herein within the device, and may be embodied in any suitable form and organized within any suitable type of data structure. The operational data may be collected as a single data set, or may be distributed over different locations including over different storage devices, and may exist, at least partially, as electronic signals on a system or network[14].

Reference throughout this specification to “a select embodiment,” “one embodiment,” or “an embodiment” means that a particular feature, structure, or characteristic described in connection withthe embodiment is included in at least one embodiment of the disclosed subject matter. Thus, appearances of the phrases “a select embodiment,” “in one embodiment,” or “in an embodiment” in various places throughout this specification are not necessarily referring to the same embodiment[15].

Furthermore, the described features, structures, or characteristics may be combined in any suitable manner in one or more embodiments. In the following description, numerous specific details are provided, to provide a thorough understanding of embodiments of the disclosed subject matter. One skilled in the relevant art will recognize, however, that the disclosed subject matter can be practiced without one or more of the specific details, or with other methods, components, materials, etc. In other instances, well-known structures, materials, or operations are not shown or described in detail toavoid obscuring aspects of the disclosed subject matter[16].

In accordance with the exemplary embodiments, the disclosed computer programs or modules can beexecuted in many exemplary ways, such as an application that is resident in the memory of a device oras a hosted application that is being executed on a server and communicating with the device application or browser via a number of

standard protocols, such as TCP/IP, HTTP, XML, SOAP, REST, JSON and other sufficient protocols. The disclosed computer programs can be written in exemplary programming languages that execute from memory on the device or from a hosted server, such as BASIC, COBOL, C, C++, Java, Pascal, or scripting languages such as JavaScript, Python, Ruby, PHP, Perl or other sufficient programming languages[17].

Some of the disclosed embodiments include or otherwise involve data transfer over a network, such ascommunicating various inputs or files over the network. The network may include, for example, one or more of the Internet, Wide Area Networks (WANs), Local Area Networks (LANs), analog or digital wired and wireless telephone networks (e.g., a PSTN, Integrated Services Digital Network (ISDN), a cellular network, and Digital Subscriber Line (xDSL)), radio, television, cable, satellite, and/or any other delivery or tunneling mechanism for carrying data. The network may include multiple networks or sub networks, each of which may include, for example, a wired or wireless data pathway. The network may include a circuit-switched voice network, a packet-switched data network, or any other network able to carry electronic communications. For example, the network may include networks based on the Internet protocol (IP) or asynchronous transfer mode (ATM), andmay support voice using, for example, VoIP, Voice-over-ATM, or other comparable protocols used for voice data communications. In one implementation, the network includes a cellular telephone network configured to enable exchange of text or SMS messages[18].

Examples of the network include, but are not limited to, a personal area network (PAN), a storage area network (SAN), a home area network (HAN), a campus area network (CAN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), a virtual private network (VPN), an enterprise private network (EPN), Internet, a global area network (GAN), and so forth. The present invention will be further described in detail with reference to the accompanying drawings[19].

The present invention provides a novel method for identifying herbal compounds with anti-cancer properties using machine learning techniques. The method's advantages include increased efficiency, improved accuracy, cost-effectiveness, and comprehensive screening capabilities. The detailed description, along with the accompanying flowchart, elucidates the steps involved in implementing the invention. By leveraging the power of machine learning algorithms and comprehensive datasets, this method accelerates the identification of potential therapeutic candidates from herbal compounds,thereby contributing to advancements in anti-cancer drug discovery[20].

Figure 1 illustrates a flowchart outlining the steps involved in the method for identifying herbal compounds with anti-cancer properties using machine learning techniques. The following descriptionwill provide a more detailed explanation of each step[21].

Step 1 (102): Data Collection In this step, a comprehensive dataset is compiled, incorporating information on herbal compounds and their chemical properties, as well as existing experimental data on their anti-cancer activities. This data can be obtained from various sources, including scientific literature, public databases, and experimental studies.

Step 2 (104): Feature Extraction The compiled dataset is subjected to feature extraction, where relevant features are identified and extracted. These features may include molecular descriptors, physicochemical properties, structural characteristics, and any other relevant information that can describe the herbal compounds and their potential anti-cancer properties. Feature selection techniques can be applied to reduce the dimensionality of the dataset and eliminate irrelevant or redundant features[22].

Step 3 (106): Data Preprocessing In this step, the dataset undergoes preprocessing to ensure data quality and

consistency. Noise is removed, and missing values are handled through appropriate imputation techniques. Furthermore, the features are normalized or scaled to a common range to prevent biases caused by differences in their scales[23].

Step 4 (108): Model Training Machine learning models, such as support vector machines, random forests, or neural networks, are employed to train on the preprocessed dataset. The models learn fromthe labeled examples in the dataset, capturing the complex relationships between the extracted

features and the corresponding anti-cancer activities of the herbal compounds. The training process involves optimization techniques, such as gradient descent, to iteratively update the model parameters and minimize the prediction error[23].

Step 5 (110): Model Validation To assess the performance and generalizability of the trained models, cross-validation techniques are employed. The dataset is divided into multiple subsets, and the models are tested on different combinations of training and validation sets. Evaluation metrics, such as accuracy, precision, recall, and area under the curve (AUC), are calculated to measure the models' performance in predicting the anti-cancer activities of the herbal compounds accurately[24].

Step 6 (112): Prediction and Compound Selection Once the models are validated, they are ready topredict the anti-cancer activities of new, untested herbal compounds. The extracted features of these compounds are inputted into the trained models, which generate predictions based on the learnedpatterns. The predicted activities serve as a basis for ranking and prioritizing the herbal compounds. According to their potential efficacy as anti-cancer agents. The compounds with the highest predicted activities are selected for further experimental validation.

Step 7 (114): Experimental Validation In this final step, the top-ranked herbal compounds selected based on the predictions undergo experimental validation to confirm their anti-cancer properties. Various assays and tests, such as cell viability assays, apoptosis assays, and animal studies, can be conducted to assess the compounds' effectiveness in inhibiting cancer growth or inducing cancer cell death. The results obtained from the experimental validation further refine the selection of herbal compounds with significant anti-cancer activity[25].
The method for identifying herbal compounds with anti-cancer properties using machine learning techniques described herein offers a more efficient and accurate approach compared to conventional screening methods. By leveraging the power of machine learning algorithms and comprehensive datasets, the invention enables the rapid screening and selection of potential therapeutic candidates from the vast pool of herbal compounds, contributing to the advancement of anti-cancer drug discovery.
The method involves the following steps:

Data Collection: A comprehensive dataset is compiled, consisting of information on herbal compounds, their chemical properties, and existing experimental data on their anti-cancer activities.
Feature Extraction: Relevant features are extracted from the dataset, including molecular descriptors, physicochemical properties, and structural characteristics of the herbal compounds[26].

Data Preprocessing: The dataset is preprocessed to remove noise, handle missing values, and normalize the features to ensure consistency.

Model Training: Machine learning models, such as support vector machines, random forests, or neuralnetworks, are

trained using the preprocessed dataset. These models learn the complex relationships between the chemical features and the anti-cancer activities of the herbal compounds.

Model Validation: The trained models are evaluated using cross-validation techniques to assess their performance and generalizability.

Prediction and Compound Selection: Once the models are validated, they are applied to predict the anti-cancer activities of new, untested herbal compounds. The predicted activities are used to rank andprioritize the compounds based on their potential efficacy[27].

Experimental Validation: The top-ranked herbal compounds are selected for further experimental validation to confirm their anti-cancer properties.

The inventive method harnesses the power of machine learning algorithms to learn from the vast amount of available data, thereby enabling the identification of potential anti-cancer herbal compounds more efficiently and accurately.

In an embodiment, a system for identifying herbal compounds with anti-cancer properties comprises: a data collection module configured to collect a dataset comprising information on herbal compounds, their chemical properties, and experimental data on their anti-cancer activities; a feature extraction module configured to extract relevant features from the dataset, including molecular descriptors, physicochemical properties, and structural characteristics of the herbal compounds; a data preprocessing module configured to preprocess the dataset by removing noise, handling missing values, and normalizing the features; a machine learning module configured to train machine learning models using the preprocessed dataset to learn the relationships between the features and the anti- cancer activities of the herbal compounds. e. a validation module configured to validate the trained models using cross-validation techniques to assess their performance and generalizability; a predictionmodule configured to predict the anti-cancer activities of new, untested herbal compounds using the validated models and rank the compounds based on their predicted efficacy; and an experimental validation module configured to experimentally validate the top-ranked herbal compounds to confirmtheir anti-cancer properties[28].

In an embodiment, the system further comprises a user interface module configured to interact with users and display the results of the predicted anti-cancer activities and the ranking of herbal compounds[29].

In an embodiment, the system is implemented in a cloud computing environment, allowing for scalable and distributed processing of the dataset and prediction of anti-cancer activities for multiple herbal compounds simultaneously.

The present invention offers several advantages over conventional methods for identifying herbal compounds with anti-cancer properties:

Increased Efficiency: By utilizing machine learning techniques, the method accelerates the screening process, allowing for the rapid identification of potential therapeutic candidates[29].

Improved Accuracy: The predictive models developed through machine learning algorithms considerthe complex interactions between herbal compounds and their targets, resulting in more accurate predictions.

Cost-effectiveness: The method reduces the reliance on extensive experimental testing, saving resources and minimizing costs associated with traditional screening methods.

Comprehensive Screening: The method enables the simultaneous screening of a large number of herbal

compounds, facilitating the identification of compounds with diverse anti-cancer mechanismsand broad therapeutic potential[30].

In an embodiment, the below herbal compounds having anti-cancer properties can be included in thedataset used in the present invention:

Curcumin: Derived from the turmeric plant, curcumin has demonstrated anti-cancer effects by inhibiting tumor growth, inducing apoptosis (programmed cell death), and suppressing inflammation. It has shown potential in various cancer types, including breast, colorectal, lung, and pancreatic cancer.

Resveratrol: Found in grapes, red wine, and berries, resveratrol possesses anti-cancer properties through its antioxidant and anti-inflammatory actions. It has been investigated for its potential in preventing and treating various cancers, including breast, prostate, and colorectal cancer[31].

Epigallocatechin gallate (EGCG): Found abundantly in green tea, EGCG has shown promising anti- cancer effects by inhibiting tumor cell proliferation, inducing apoptosis, and suppressing angiogenesis. It has been studied for its potential in preventing and treating breast, lung, and prostate cancer.

Quercetin: Widely present in fruits, vegetables, and herbs, quercetin exhibits anti-cancer properties through its antioxidant and anti-inflammatory actions. It has been investigated for its potential in various cancer types, including lung, breast, and colon cancer.

Astragalus polysaccharides:
Derived from the Astragalus plant, these polysaccharides have demonstrated anti-cancer effects by enhancing immune function, inhibiting tumor cell proliferation, and inducing apoptosis. They have been studied for their potential in treating liver, lung, and gastric cancer[32].
These compounds serve as valuable references for training the models to recognize patterns and associations between chemical features and anti-cancer activities.
The drawings and the forgoing description give examples of embodiments. Those skilled in the art will appreciate that one or more of the described elements may well be combined into a single functional element. Alternatively, certain elements may be split into multiple functional elements. Elements from one embodiment may be added to another embodiment. For example, orders of processes described herein may be changed and are not limited to the manner described herein[33].
Moreover, the actions of any flow diagram need not be implemented in the order shown; nor do all of the acts necessarily need to be performed. Also, those acts that are not dependent on other acts may be performed in parallel with the other acts.
The scope of embodiments is by no means limited by these specific examples. Numerous variations, whether explicitly given in the specification or not, such as differences in structure, dimension, and use of material, are possible. The scope of embodiments is atleast as broad as given by the following claims[34].
Benefits, other advantages, and solutions to problems have been described above with regard to specific embodiments. However, the benefits, advantages, solutions to problems, and any component(s) that may cause any benefit, advantage, or solution to occur or become more pronouncedare not to be construed as a critical, required, or essential feature or component of any or all the claims.

Results and discussion

A method for identifying herbal compounds with anti-cancer properties using machine learning techniques, comprising:collecting a dataset comprising information on herbal compounds, their chemicalproperties, and experimental data on their anti-cancer activities; Extracting relevant features from the dataset, including molecular descriptors,physicochemical properties, and structural characteristics of the herbal compounds[35]; Preprocessing the dataset by removing noise, handling missing values, and normalizing thefeatures; Training machine learning models using the preprocessed dataset to learn the relationships between the features and the anti-cancer activities of the herbal compounds; Validating the trained models using cross-validation techniques to assess their performance and generalizability; Predicting the anti-cancer activities of new, untested herbal compounds using the validated models and ranking the compounds based on their predicted efficacy; and Experimentally validating the top-ranked herbal compounds to confirm their anti-cancerproperties. The method as described wherein the machine learning models comprise support vectormachines, random forests, or neural networks. The method as described wherein the dataset is collected from scientific literature, publicdatabases, and experimental studies[36]. The method as described wherein the features are selected using feature selection techniques to reduce the dimensionality of the dataset. The method as described wherein the dataset is preprocessed by imputing missing values usingappropriate techniques and normalizing the features to a common range.

The method as described wherein the performance of the trained models is evaluated using accuracy, precision, recall, or area under the curve (AUC) metrics.

The method as described wherein the experimental validation comprises cell viability assays, apoptosis assays, or animal studies to assess the anti-cancer properties of the selected herbal compounds. A system for identifying herbal compounds with anti-cancer properties, comprising: a data collection module configured to collect a dataset comprising information on herbal compounds, their chemical properties, and experimental data on their anti-cancer activities; a feature extraction module configured to extract relevant features from the dataset, including molecular descriptors, physicochemical properties, and structural characteristics of the herbal compounds; a data preprocessing module configured to preprocess the dataset by removing noise, handling missing values, and normalizing the features[38];

A machine learning module configured to train machine learning models using thepreprocessed dataset to learn the relationships between the features and the anti-cancer activities of the herbal compounds. e. a validation module configured to validate the trained models using cross-validation techniques to assess their performance and generalizability; a prediction module configured to predict the anti-cancer activities of new, untested herbal compounds using the validated models and rank the compounds based on their predicted efficacy; and an experimental validation module configured to experimentally validate the top-ranked herbal compounds to confirm their anti-cancer properties. The system as described, further comprising a user interface module configured to interact with users and display the results of the predicted anti-cancer activities and the ranking of herbal compounds. The system as described, wherein the system is implemented in a cloud computingenvironment, allowing for scalable and distributed processing of the dataset and prediction of anti-cancer activities for multiple herbal compounds simultaneously[39-40].

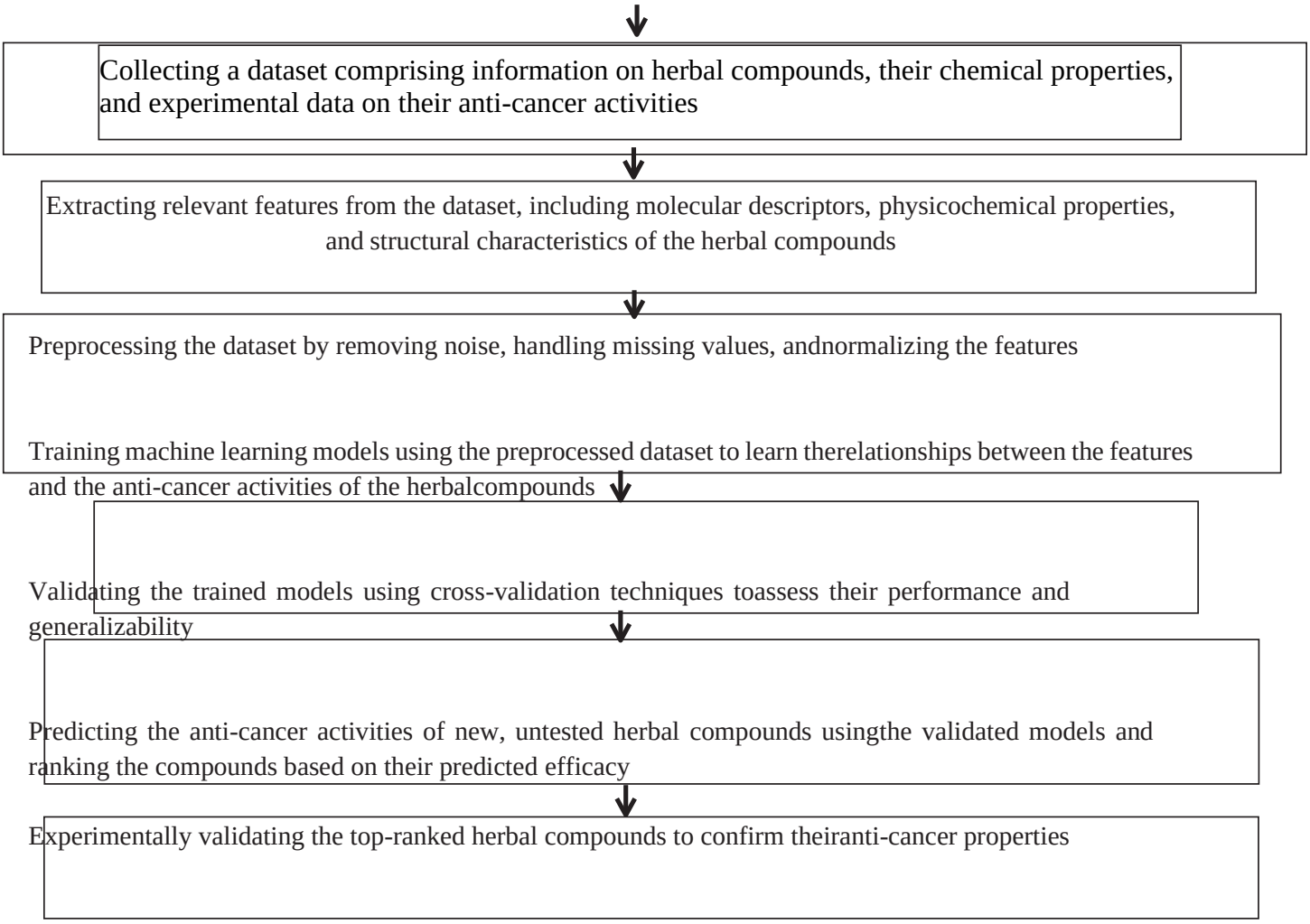

Figure 1: Illustrates Flowchart illustrating the steps involved in the method for identifying herbal compounds with anti-cancer properties using machine learning techniques.

References

1. Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18(5):435–41. [PMC free article] [PubMed] [Google Scholar]
2. Nosengo N Can you teach old drugs new tricks? Nature. 2016;534(7607):314–6. [PubMed] [Google Scholar]
3. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3(11):935–49. [PubMed] [Google Scholar]
4. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47(7):1739–49. [PubMed] [Google Scholar]
5. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61. [PMC free article] [PubMed] [Google Scholar]
6. Jain AN. Scoring functions for protein-ligand docking. Curr Protein Pept Sci. 2006;7(5):407–20. [PubMed] [Google Scholar]

7. Wang Y, Lin HQ, Wang P, Hu JS, Ip TM, Yang LM, et al. Discovery of a Novel HIV-1 Integrase/p75 Interacting Inhibitor by Docking Screening, Biochemical Assay, and in Vitro Studies. J Chem Inf Model. 2017;57(9):2336–43. [PubMed] [Google Scholar]
8. Mirza SB, Salmas RE, Fatmi MQ, Durdagi S. Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study. J Mol Graph Model. 2016;66:99–107. [PubMed] [Google Scholar]
9. Kumar V, Krishna S, Siddiqi MI. Virtual screening strategies: recent advances in the identification and design of anti-cancer agents. Methods. 2015;71:64–70. [PubMed] [Google Scholar]
10. Hafeez BB, Ganju A, Sikander M, Kashyap VK, Hafeez ZB, Chauhan N, et al. Ormeloxifene Suppresses Prostate Tumor Growth and Metastatic Phenotypes via Inhibition of Oncogenic beta-catenin Signaling and EMT Progression. Mol Cancer Ther. 2017;16(10):2267–80. [PMC free article] [PubMed] [Google Scholar]
11. Chen YC. Beware of docking! Trends Pharmacol Sci. 2015;36(2):78–95. [PubMed] [Google Scholar]
12. Wallach I, Dzamba M, Heifets A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. arXiv e-prints [Internet]. 2015. October 01, 2015. Available from: https://ui.adsabs.harvard.edu/abs/2015arXiv151002855W. [Google Scholar]
13. Desaphy J, Bret G, Rognan D, Kellenberger E. sc-PDB: a 3D-database of ligandable binding sites--10 years on. Nucleic Acids Res. 2015;43(Database issue):D399–404. [PMC free article] [PubMed] [Google Scholar]
14. Spitzer R, Jain AN. Surflex-Dock: Docking benchmarks and real-world application. J Comput Aided Mol Des. 2012;26(6):687–99. [PMC free article] [PubMed] [Google Scholar]
15. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, et al. DOCK 6: Impact of new features and current docking performance. J Comput Chem. 2015;36(15):1132–56. [PMC free article] [PubMed] [Google Scholar]
16. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein-Ligand Scoring with Convolutional Neural Networks. J Chem Inf Model. 2017;57(4):942–57. [PMC free article] [PubMed] [Google Scholar]
17. Li Y, Han L, Liu Z, Wang R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. J Chem Inf Model. 2014;54(6):1717–36. [PubMed] [Google Scholar]
18. Nguyen DD, Cang Z, Wu K, Wang M, Cao Y, Wei GW. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. J Comput Aided Mol Des. 2019;33(1):71–82. [PMC free article] [PubMed] [Google Scholar]
19. Maldonado AG, Doucet JP, Petitjean M, Fan BT. Molecular similarity and diversity in chemoinformatics: from theory to applications. Mol Divers. 2006;10(1):39–79. [PubMed] [Google Scholar]
20. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462(7270):175–81. [PMC free article] [PubMed] [Google Scholar]
21. Hu Y, Stumpfe D, Bajorath J. Advancing the activity cliff concept. F1000Res. 2013;2:199. [PMC free article] [PubMed] [Google Scholar]
22. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 2016;44(D1):D1045–53. [PMC free article] [PubMed] [Google Scholar]
23. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019;47(D1):D1102–D9. [PMC free article] [PubMed] [Google Scholar]
24. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 2018;10(4):e1417. [PMC free article] [PubMed] [Google Scholar]
25. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945–D54. [PMC free article] [PubMed] [Google Scholar]

26. Deshmukh AL, Chandra S, Singh DK, Siddiqi MI, Banerjee D. Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening. Mol Biosyst. 2017;13(8):1630–9. [PubMed] [Google Scholar]
27. Algamal ZY, Lee MH, Al-Fakih AM, Aziz M. High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO. Journal of Chemometrics. 2015;29(10):547–56. [Google Scholar]
28. Alam S, Khan F. 3D-QSAR studies on Maslinic acid analogs for Anticancer activity against Breast Cancer cell line MCF-7. Sci Rep. 2017;7(1):6019. [PMC free article] [PubMed] [Google Scholar]
29. Sterling T, Irwin JJ. ZINC 15--Ligand Discovery for Everyone. J Chem Inf Model. 2015;55(11):2324–37. [PMC free article] [PubMed] [Google Scholar]
30. Taha MO, Al-Sha'er MA, Khanfar MA, Al-Nadaf AH. Discovery of nanomolar phosphoinositide 3-kinase gamma (PI3Kgamma) inhibitors using ligand-based modeling and virtual screening followed by in vitro analysis. Eur J Med Chem. 2014;84:454–65. [PubMed] [Google Scholar]
31. Allen BK, Ayad NG, Schürer SC. Kinome-wide activity classification of small molecules by deep learning. bioRxiv. 2019:512459. [Google Scholar]
32. Schurer SC, Muskal SM. Kinome-wide activity modeling from diverse public high-quality data sets. J Chem Inf Model. 2013;53(1):27–38. [PMC free article] [PubMed] [Google Scholar]
33. Rifaioglu AS, Atalay V, Martin MJ, Cetin-Atalay R, Doğan T. DEEPScreen: High Performance Drug-Target Interaction Prediction with Convolutional Neural Networks Using 2-D Structural Compound Representations. bioRxiv. 2018:491365. [PMC free article] [PubMed] [Google Scholar]
34. Cho WC. OncomiRs: the discovery and progress of microRNAs in cancers. Mol Cancer. 2007;6:60. [PMC free article] [PubMed] [Google Scholar]
35. Jamal S, Periwal V, Open Source Drug Discovery C, Scaria V. Computational analysis and predictive modeling of small molecule modulators of microRNA. J Cheminform. 2012;4(1):16. [PMC free article] [PubMed] [Google Scholar]
36. Liu X, Wang S, Meng F, Wang J, Zhang Y, Dai E, et al. SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. Bioinformatics. 2013;29(3):409–11. [PubMed] [Google Scholar]
37. Wang CC, Chen X, Qu J, Sun YZ, Li JQ. RFSMMA: A New Computational Model to Identify and Prioritize Potential Small Molecule-MiRNA Associations. J Chem Inf Model. 2019;59(4):1668–79. [PubMed] [Google Scholar]
38. Qu J, Chen X, Sun YZ, Zhao Y, Cai SB, Ming Z, et al. In Silico Prediction of Small Molecule-miRNA Associations Based on the HeteSim Algorithm. Mol Ther Nucleic Acids. 2019;14:274–86. [PMC free article] [PubMed] [Google Scholar]
39. Chen X, Guan NN, Sun YZ, Li JQ, Qu J. MicroRNA-small molecule association identification: from experimental results to computational models. Brief Bioinform. 2018. [PubMed] [Google Scholar]
40. Gessi S, Merighi S, Sacchetto V, Simioni C, Borea PA. Adenosine receptors and cancer. Biochim Biophys Acta. 2011;1808(5):1400–12. [PubMed] [Google Scholar]