



Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Review

Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine



Vivek Bhakta Mathema ^{a,b,1}, Partho Sen ^{c,d}, Santosh Lamichhane ^c, Matej Orešić ^{c,d},
Sakda Khoomrung ^{a,b,e,*,2}

^a Metabolomics and Systems Biology, Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^b Siriraj Metabolomics and Phenomics Center, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^c Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland

^d School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden

^e Center of Excellence for Innovation in Chemistry (PERCH-CIC), Faculty of Science, Mahidol University, Bangkok, Thailand

ARTICLE INFO

Article history:

Received 17 August 2022

Received in revised form 28 January 2023

Accepted 29 January 2023

Available online 31 January 2023

Keywords:

Cancer
Deep learning
Reinforcement learning
Precision medicine
Oncogene
Systems medicine

ABSTRACT

Cancer progression is linked to gene-environment interactions that alter cellular homeostasis. The use of biomarkers as early indicators of disease manifestation and progression can substantially improve diagnosis and treatment. Large omics datasets generated by high-throughput profiling technologies, such as microarrays, RNA sequencing, whole-genome shotgun sequencing, nuclear magnetic resonance, and mass spectrometry, have enabled data-driven biomarker discoveries. The identification of differentially expressed traits as molecular markers has traditionally relied on statistical techniques that are often limited to linear parametric modeling. The heterogeneity, epigenetic changes, and high degree of polymorphism observed in oncogenes demand biomarker-assisted personalized medication schemes. Deep learning (DL), a major subunit of machine learning (ML), has been increasingly utilized in recent years to investigate various diseases. The combination of ML/DL approaches for performance optimization across multi-omics datasets produces robust ensemble-learning prediction models, which are becoming useful in precision medicine. This review focuses on the recent development of ML/DL methods to provide integrative solutions in discovering cancer-related biomarkers, and their utilization in precision medicine.

© 2023 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	1373
2. Deep learning-assisted medical imaging and multi-omics data analysis in cancer research for early cancer diagnosis	1374
3. <i>De novo</i> synthesis and prediction of compound properties and activities using deep learning approach	1375
4. Potentials of deep learning in tumor-associated metabolic pathways prediction and graph-based convolution neural networks in cancer studies	1376
5. Deep learning-based ensemble modeling for precision cancer medicine	1378
6. Combination of deep learning and deep reinforcement learning	1378
7. Summary and Outlook	1379
CRedit authorship contribution statement	1380
Declarations of interest	1380
Acknowledgements	1380

* Corresponding author at: Metabolomics and Systems Biology, Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.
E-mail address: sakda.kho@mahidol.edu (S. Khoomrung).

¹ <https://orcid.org/0000-0003-3916-9949>

² <https://orcid.org/0000-0001-9461-8597>

References	1380
Glossary	1382

1. Introduction

Molecular biomarkers are physiological indicators that can indicate illness-associated alterations at the molecular level, assist in the prognosis of disease manifestation, and identify disease-related molecular targets [1–3]. In cancer pathology, the use of appropriate biomarkers for early diagnosis and prognosis is crucial to reducing mortality. Early diagnosis and prognosis of cancer are complicated by genetic polymorphisms, the presence of oncogenes, and epigenetic variables [4,5]. Data integration strategies have aided patient clinical management in recent years by improving diagnostic accuracy and therapeutic efficacy [5]. In contrast to the intelligence of animals and people, artificial intelligence (AI) is the intelligence of machines that can perceive, synthesize, and infer knowledge. Machine learning (ML) is a kind of AI that can accurately predict outcomes based on training data without being specifically pre-configured to do so (Fig. 1a). The invention of the artificial neural network (ANN) allowed the modeling of complex non-linear systems by depicting the mechanisms of biological neurons. The ANN is made up of a network of connected units or nodes known as artificial neurons (Fig. 1b), which are modeled after neurons in the human brain. Each connection, similar to synapses in the human brain has the ability to transmit a signal to other nodes. An artificial neuron receives, analyzes, and communicates with other connected neurons. The output of each neuron is determined by some non-linear function of the sum of its inputs, and the "signal" is sent as a real number output. In particular, deep learning (DL) is a subset of ML approaches (Fig. 1a) that combine ANNs and representation learning. The learning can allow for highly dense and fully connected multi-layered networks that can be trained in supervised, semi-

supervised, or unsupervised settings (Fig. 1c). These ANNs can also be used to construct autoencoders that use unsupervised learning techniques for data encoding.

Conventional polygenic risk scoring through the utilization of AI and ML is also becoming a promising tool for the early detection and prognosis of cancer [6,7]. In this context, deep learning may provide a suitable alternative for modeling complex traits and incorporating previously restricted multidimensional medical imaging datasets [8–11]. Although the conventional ML approach is the most efficient analysis technique in multiple medical discovery and clinical decision support systems (CDSS), the utility of DL is gaining attention in multi-data-type analysis involving ensemble-based disease study models [5,12,13]. Furthermore, DL-assisted identification of susceptible genes and associated proteomics and metabolomics profiles could be a useful strategy to detect cancers at early stages [5,14–18]. Such an integrative strategy may also provide targets for precision medicine, which may help to enhance the prognosis for total recovery. Recent studies suggest that a DL-based approach that integrates a broad range of datasets, including histology, magnetic resonance imaging (MRI), X-rays, and chromatograms, along with multi-omics data can significantly improve the accuracy of diagnostic models for cancer [8,19,20]. Precision oncology requires the ability to predict outcomes for specific cancer patients, such as survival or metastasis [21]. DL approaches, such as graph neural networks (GNN), applied to gene regulation and metabolic pathways, are emerging as useful tools in the study of tumor metastasis [21–23]. GNN is a form of neural network that acts directly on the graph structure consisting of multiple nodes, each representing an entity. Node categorization is a common utility of GNN. Moreover, the development of generative DL models is currently being utilized

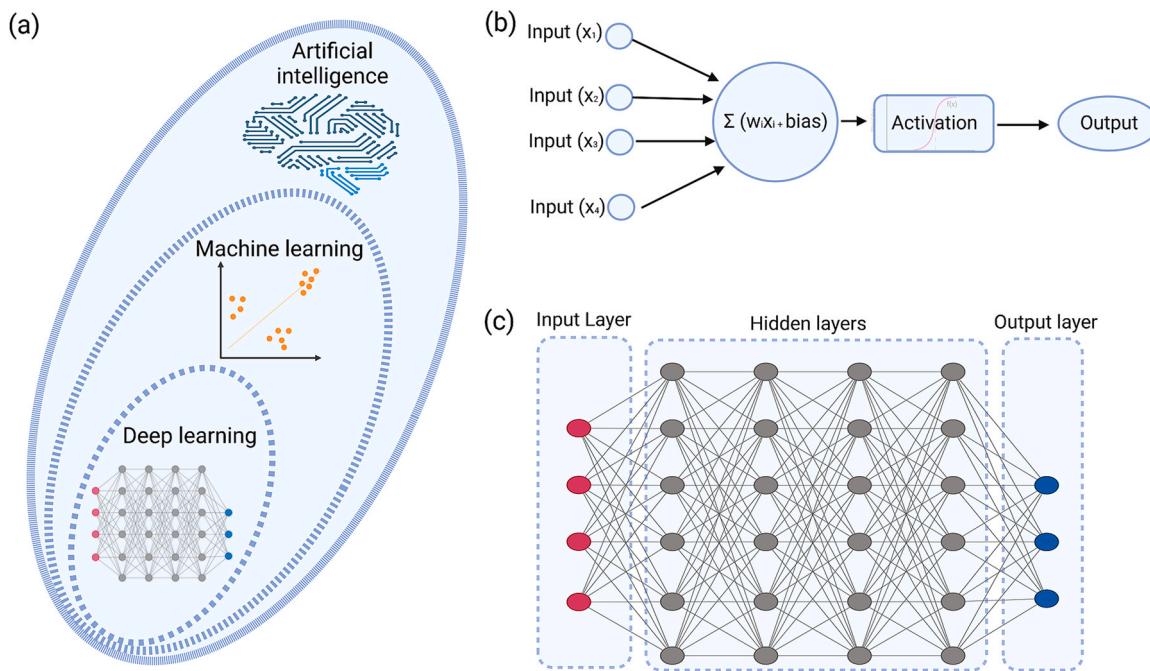


Fig. 1. Basics of deep learning models. (a) Deep learning is a subset of the machine learning approach, which itself is a subdomain of artificial intelligence. The amount of dots in boundary line schematically represents amount of datasets available for model training (b) A basic unit of artificial neural network is a neuron, which can take input and undergo calculations involving weight and biases. The resulting values are subjected to an activation function to decide whether or not the neuron's input to the network is significant in the prediction process utilizing simpler mathematical procedures. (c) Deep learning models consist of multiple layers of such individual neurons with the ability to process high-dimension data. The dataset is fed into the network in the input layer, which is subsequently followed by multiple inner hidden layers, and final outcomes are computed on the outer layer with the utility of different activation functions.

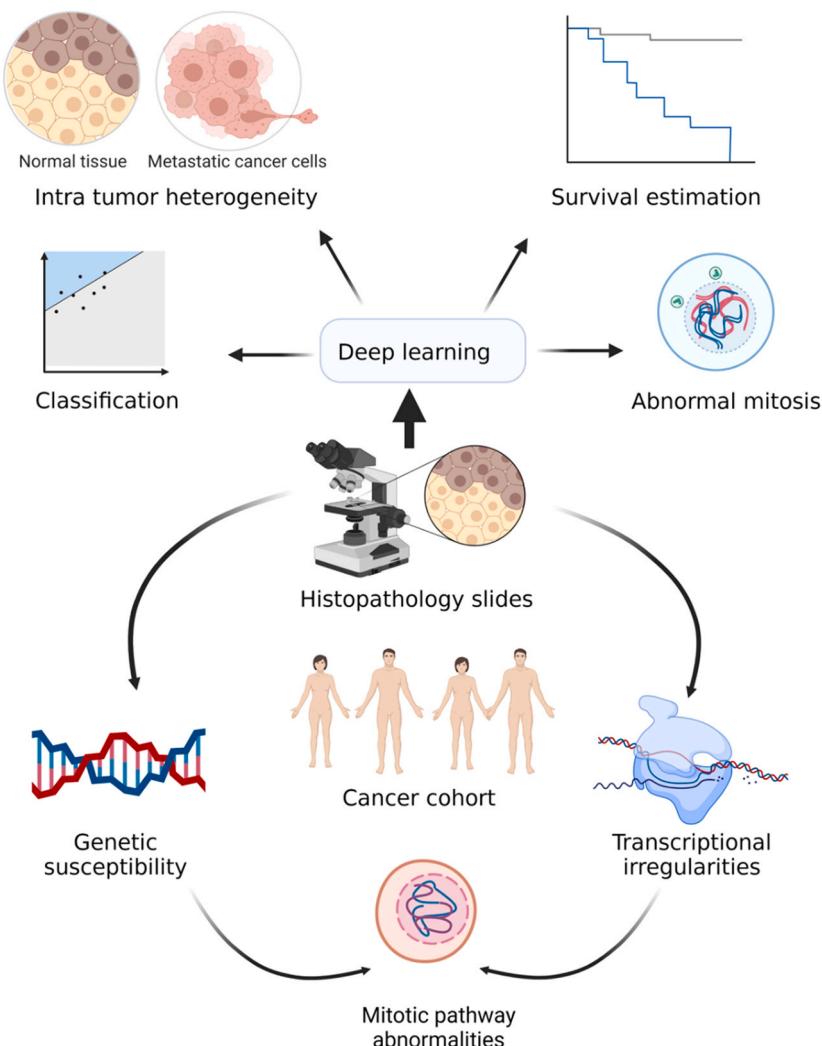


Fig. 2. Integration of histological slides and genetic susceptibility data in deep learning techniques for malignancy prediction. Histological slides provide strong evidence related to clinical manifestations of cancer such as neoplasms, malignant tumors, and metastasis. Deep learning techniques can combine these traditional image-based datasets with well-known genomic tests to make a strong model for early cancer diagnosis that is much more accurate than individual tests.

for drug target discovery and *de novo* synthesis of investigational new drugs to facilitate cancer research [24,25]. This review briefly explores the current trends in DL, which facilitate multi-data type analyses, biomarker discoveries, and precision medicine.

2. Deep learning-assisted medical imaging and multi-omics data analysis in cancer research for early cancer diagnosis

Medical image categorization is essential in clinical diagnosis, prognosis, and educational activities [26,27]. Imaging biomarkers are based on the anatomical, histological, or radiographic characteristics of samples. Histology slides are a valuable source in clinical practice for screening cancer biomarkers such as angiogenesis, tumor growth, and metastasis [28,29]. Manual evaluation of histological slides (Fig. 2), X-ray, computed tomography (CT), and MRI are laborious and often subject to human error, leading to misdiagnosis. In recent years, DL has demonstrated exceptional accuracy in processing medical imaging data for disease diagnosis, including chest radiography, breast cancer screenings (or mammograms), and CT scans [8,18,30,31]. The success of DL in medical imaging is driven mainly by the development of convolutional neural networks (CNNs), a class of deep neural networks (DNNs) that perform

excellent visual imagery feature detection with high sensitivity and specificity [32]. 3D versions of CNN also assist in solving problems related to 3D image segmentation, which is necessary for categorizing brain tissues, heart structures, and abdominal organs [8,33]. The transfer learning (TL) is a DL-based strategy that focuses on retaining information learned while addressing one issue and transferring it to resolve a similar but distinct problem. Well-established CNNs such as VGG16 [34] which have been trained to recognize a variety of objects, can be leveraged through TL and applied to relatively small datasets to yield highly accurate diagnosis models [35,36]. Similarly, one-shot and few-shot learning models are also being used for semantic feature learning tasks to help diagnose lung cancer recurrences [37] and lymph node metastasis [38,39]. However, the datasets used for training these models have a significant impact on their accuracy. Therefore, the quality of training datasets can highly influence the performance of models [40]. CNN can also detect features in non-image data such as gene expression via transforming non-image datasets into feature-rich image vectors. This can be subjected to a trained CNN classifier for cancer-type prediction based on gene expression data [41,42]. Recent developments in DL include the swarm learning (SL) technique, which involves parties working together to build AI models, while avoiding

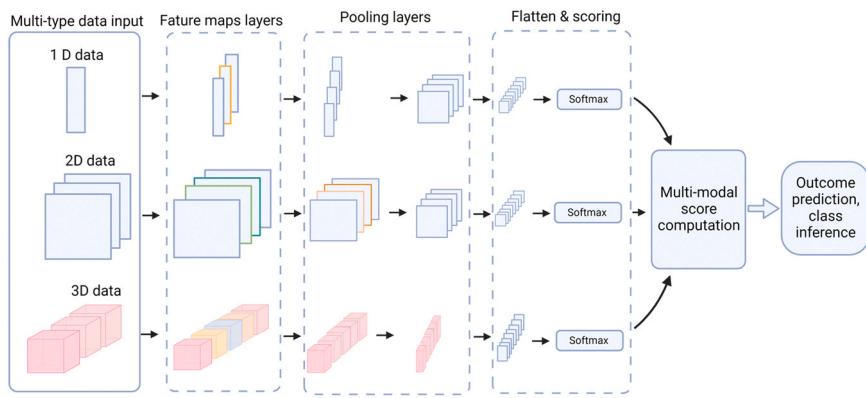


Fig. 3. Schematics illustrating utility of simple multi-modal deep learning approach in cancer research. The multi-type data input can include blood metabolite concentrations (1D data), histological or X-ray images (2D data), and MRI or CT scan images (3D data). The data undergo feature mapping followed by multiple pooling operations within DNN. The layers are subsequently flattened, and a score is computed using a scoring function (e.g. softmax) for each model. These scores from multiple models are further processed to compute the final outcome as disease class prediction or inference.

data transfer and monopolistic data governance for large multicentric datasets. Swarm learning was implemented on multi-centric datasets of histopathological images from over 5000 cases. It assisted in the prediction of BRAF protein mutation status and microsatellite instability directly from colorectal cancer histological slides [43]. Together, these studies suggest that DL is paving the way to improve early cancer diagnosis using medical images. Independent dataset validation is a commonly used technique in AL and DL for evaluating the performance of a trained model. This technique involves using a separate dataset for evaluation, which is isolated from the training and testing dataset, to ensure the generalization of model to unseen data. Independent dataset validation have been utilized to evaluate the performance of several DL models ranging from imaging biomarkers in neurodegenerative such as Alzheimer [44,45], breast cancer [18], histopathological analysis for malaria [46] to time series-labelled tumor markers [47]. DL is an emerging field, thus, it is important to note that due to the lack of standardized protocols for data collection and the limitations in technical capabilities, not all models or studies have the ability to utilize independent dataset validation. This can pose a challenge for ensuring the generalizability of model performance and should be taken into consideration when designing and evaluating deep learning models.

DL may assist in the in-depth study of cancer by enabling the integrative analysis of multi-omics datasets [48–51]. A generalized schema (Fig. 3) involving multi-type data input illustrates the utility of multi-modal DL approach for data classification or inference. The schematics summarizes intake of hypothetical multi-dimension data such as: blood metabolite concentrations, histological images of biopsy, and CT scans to classify or inference possible outcomes of a cancer-related diagnosis. These input datasets can be of any type, such as X-ray images, blood metabolite concentrations, microarrays, and genomics data. Lee et al. integrated a DL-based autoencoding method with multi-omics datasets, including genomic, transcriptomic, and epigenomic (mRNA, miRNA, DNA methylation, and copy number variations) datasets obtained from patients with lung adenocarcinoma (LUAD). The model was able to learn representative features of the disease to differentiate the patient subgroups, showing the potential of DL in LUAD prognosis [48]. Multi-omics data associated with breast cancer have frequently been used in DL studies [18]. In a separate study, Zhang et al. applied DL methods by integrating gene expression, gene copy number variations, miRNA expression, and DNA methylation data to predict muscle-invasive bladder cancer (MIBC) [49]. This study identified the genomic and immunological differences between high- and low-risk patients. It also recognizes the role of Keratin 7 (KRT7) as a biomarker of MIBC.

Recently, Hira et al. showed that DL-based variational autoencoders could be utilized for the classification of ovarian cancer transcriptional subtypes using multi-omics data from the cancer genome atlas (TCGA) datasets with model accuracies ranging from 87% to 93% [52]. A DL approach was also used in the identification of features linked to differential survival of patients with hepatocellular carcinoma (HCC) [53]. A DL-based survival-sensitive model was developed using RNA/miRNA sequencing and DNA methylation data obtained from patients with HCC in the TCGA database. The model stratified two optimal subgroups of patients with significant survival differences and was validated using five external datasets from various omics types. This study suggests that DL can be used for the prognosis of HCC [53]. Also, Poirion et. al. suggested a hybrid approach using a novel DL and ML framework called "DeepProg" which utilized multi-omics data to build a highly predictive model for patient survival risks among various cancer subtypes [5].

The promising outcomes of DL-based studies have accelerated broader interest in combining metabolomics and other omics data for similar applications, such as potential biomarker discovery [14,53,54], chemical structure identification [55,56], and disease comorbidity analysis [57]. Although implementing a DL approach with genomic, DNA methylation, and transcriptomic data for biomarker prediction and cancer treatment selection is relatively common [54,58,59], the use of metabolomics-based biomarkers is gaining attention because of their direct relation to the cancer phenotype [15,60,61]. Nonetheless, one of the potential liabilities of using big multi-omics dataset is that it could diminish the representation of sparsely occurring diseases as the large percentage of frequently observed features in training database may create biasness in classifier models.

Furthermore, the benefits of such findings show that including multi-omics data in a DL model improves its accuracy when compared to using only a single omic dataset [48,62]. Given the growing trend of applying DL in cancer research, modeling patient outcomes and biomarker discovery may facilitate timely diagnosis and improve the clinical management of patients with cancer.

3. De novo synthesis and prediction of compound properties and activities using deep learning approach

Drugs designed using synthesis-based predictions have a long history, dating back to the early 1960 s. It is exceedingly difficult to design novel compounds using conventional tools [25]. ML/DL approaches have shown significant potential in implementing the simplified molecular-input line-entry system (SMILES) as the basis for describing molecular structures [63,64]. AI and generative

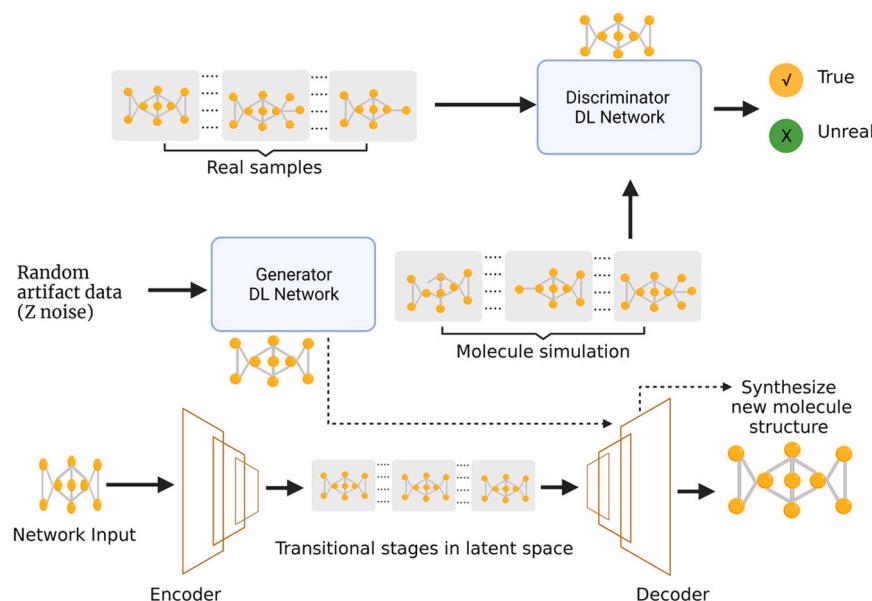


Fig. 4. Schematics of *de novo* simulation of novel compound structure based on generative deep learning models. Unlike the use of SMILES to discover potential drug structures, the deep learning-based generative models can learn molecular features of drugs in a training dataset and simulate entirely new compounds within a similar distribution. This is achieved by combining GAN with a suitable autoencoder to provide optimum results. The autoencoder is first trained using a standard molecular database as network input. The trained encoder produces latent vectors that are used as the training input for GAN using feedback obtained from the discriminator network. The new compounds are simulated by the trained generator network of GAN using random Z-noise and then utilizing the decoder to transform the sampled latent vector into a chemical structure. The solid arrow represents the flow of training data in the network. The dotted arrow represents the flow of data in the optimized GAN and autoencoder network. GAN, generative adversarial network. SMILES is a simplified molecular-input line-entry system.

models have recently been used in *de novo* drug design (DNDD) and compound optimization [25,56]. The DNDD approach is a computational growth-based method that generates new chemical structures without prior knowledge of their structural and chemical properties. The synthesis is based on the creation of new chemical entities that meet a set of requirements by employing computational growth methods. Traditional methods require ligand- or structure-based designs that are based on the active site and binder of the target molecule [65]. Other molecular generation or *de novo* design techniques include inverse QSAR10 [25,66], particle swarm optimization, and genetic algorithms [25,67]. These methods generate new molecules with specific chemical properties based on the properties of training datasets. In recent years, generative models have been increasingly used in molecular *de novo* design owing to their ability to learn the characteristics of real-world training examples and to create entirely new synthetic entities with similar characteristics [68,69]. A recent study utilized a generative adversarial network (GAN) model that integrated information from systems biology, molecular design, and transcriptome data to train a DL generative model that could autonomously design compounds with a high likelihood of imparting the desired transcriptome profile [70]. Fig. 4 shows a schematic of the recent developments in *de novo* compound design using generative models and autoencoders. Despite the lack of theoretical evidence from other ML approaches, the results suggest that a DL approach may hold potential comparable to well-established conventional ML and statistical methods in cancer studies [71,72]. Recently, promising results were reported for a *de novo* molecular generation method using a latent-vector-based GAN that did not require SMILES for compound synthesis [24]. Similarly, DL approaches have been explored to assess the feasibility of drug repurposing in oncological studies encompassing genetic profiles and drug-phenotype associations [73–75]. In association with the conventional ML approach, a range of artificial networks such as GANs, recurrent neural networks (RNNs), and autoencoders have been effectively used to generate innovative *de novo* drug creation methodologies [56,69]. Thus, DL-based generative models have the

advantage of utilizing multi-type biological training dataset and are not limited to simple molecular descriptors like SMILES. These developments highlight the potential of DL for *de novo* synthesis and compound characteristics prediction for the development of cancer drugs.

4. Potentials of deep learning in tumor-associated metabolic pathways prediction and graph-based convolution neural networks in cancer studies

A recent study by Petrovsky et al. used 1D and 3D CNNs to differentiate patients with kidney, ovarian, and prostate cancers from healthy controls by analyzing metabolomics datasets [51]. The well-trained DL model was able to recognize several oncopathologies with an area under the receiver operating characteristic curve (AUROC) ≥ 0.95 . Recently, Inglese et al. applied a DL approach to mass spectrometry (MS)-based 3D desorption electrospray ionization (DESI) imaging data to predict the heterogeneity of the metabolic phenotypes of cancer. This allowed unsupervised clustering of tumor tissue to identify subregions distinguished by the abundance of certain metabolites, revealing biological heterogeneity within the tumor [9]. In another study, DL was applied to metabolomics data to identify the metabolic features of complex traits, which accurately predicted the status of an estrogen receptor in breast cancer samples [15]. This study compared a feed-forward DL network with six ML techniques, namely random forest, support vector machine, recursive partitioning and regression trees (RPART), linear discriminant analysis, and generalized boosted models, in terms of their ability to classify negative and positive estrogen receptor breast cancer tissue samples. The DL approach achieved the highest accuracy (AUC = 0.93), outperforming well-established ML methods. Furthermore, the DL model uniquely identified eight novel metabolic pathways associated with breast cancer, paving the way for future investigation [15]. Recently, Kopylov et al. utilized a 1D CNN to recognize metabolic signatures that could distinguish between different cancer phenotypes and schizophrenia at the comorbidity level

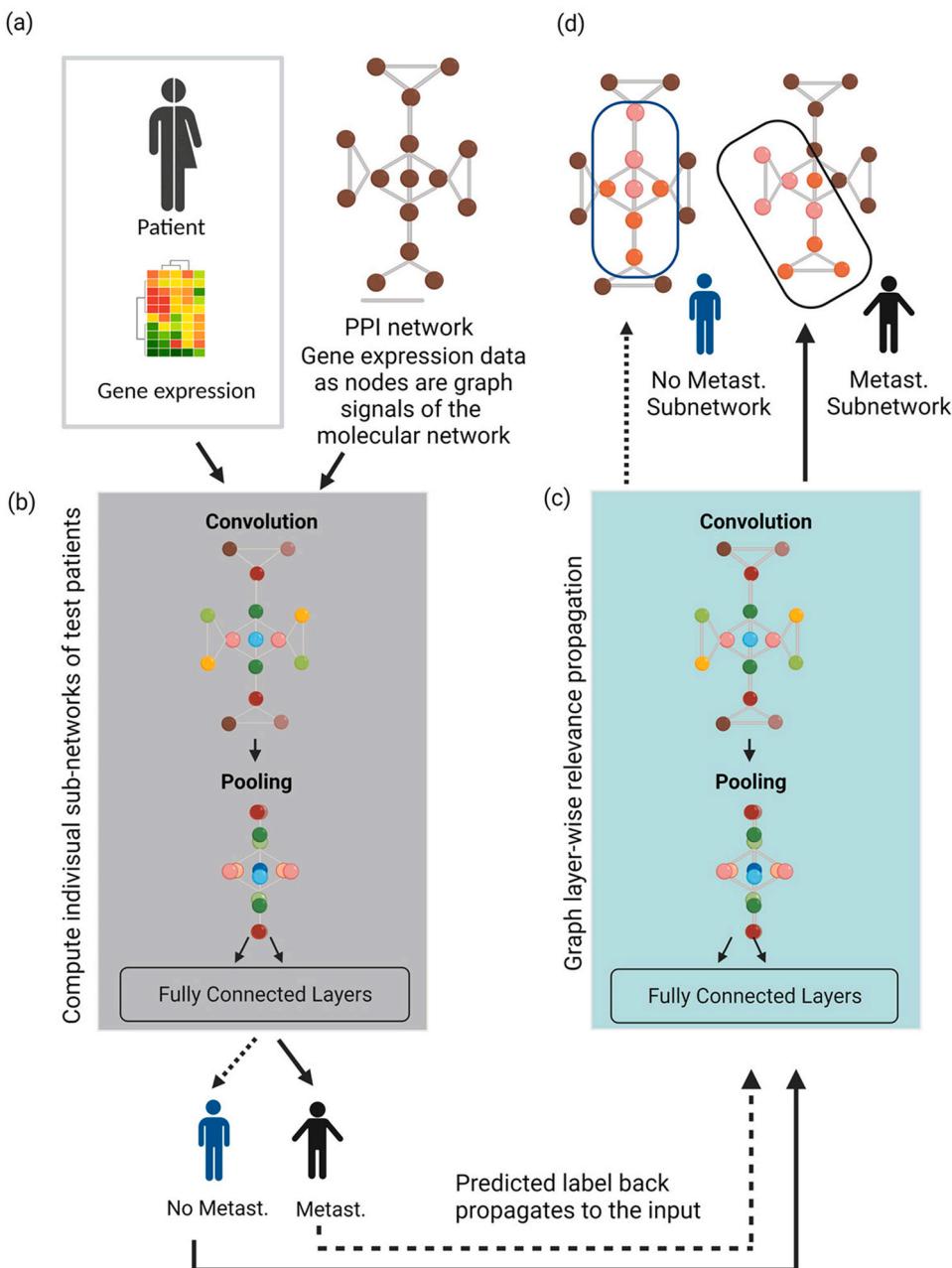


Fig. 5. Schematics of a graph convolution neural network (GCNN) for cancer diagnosis. (a) The clinical data and gene expression or proteomics PPI datasets of cancer and control groups can be parsed to (b) a specialized GCNN deep learning model, which combines this information to create a graph network-based prediction model. The GCNN uses information from training sets to customize individual patient medication regimens and (c) generate patient-specific subnetworks to (d) assist in precision medicine therapy. The blue and black oval-cornered boxes in 'd' represent metastasis and non-metastasis gene-subnetwork identified by GCNN, respectively. Each circle in 'a' represents gene expression prior to normalization or differential analysis. Multi-colored circles 'b-d' schematically represent differentially expressed genes as nodes with varying intensities. Solid and dotted arrows represent information flow in GCNN for metastatic and non-metastatic groups, respectively. GCNN, Graph convolution neural network; Metast., metastasis; PPI, protein-protein interaction network.

with a model accuracy of $\geq 90\%$ [57]. They deployed a neural network to analyze complete mass spectra and were able to efficiently distinguish unrelated pathologies and a control (healthy donors) group with a model accuracy of 90%, and different subtypes of oncophenotypes with an accuracy of 94% [57]. Zhou et al. used an autoencoder-based DL to learn the latent representative features of MS data and thereby detected paragangliomas and ovarian cancer with accuracies $> 95\%$ [76]. Shaffie et al. implemented a DL autoencoder classifier framework using CT scan images and breath-based volatile metabolic marker data as inputs to achieve over 97.8% accuracy in lung cancer diagnosis [77]. Furthermore, DL models can facilitate network inference, including metabolite–metabolite interactions

within a biochemical pathway. Fang et al. developed a DL algorithm *Lilikoi v2.0* [78] (an upgrade of *Lilikoi v1.0* [79]), which uses a multi-layer neural network trained with a stochastic gradient descent search. The *Lilikoi* algorithm can transform and incorporate metabolomics data from samples to generate personalized pathway profiles, which can aid downstream classification or development of a prognosis model based on the metabolites linked with the corresponding pathway [78]. The tool demonstrated marked alterations in the metabolic pathways associated with breast cancer in estrogen receptor-positive and -negative samples [15,78].

Indeed, the application of DL in metabolomics and lipidomics has been relatively slow, mainly because of the scarcity of reference

standards, accurate analytical methods, computational resources, expertise, and interpretability. Because DL models require large datasets for training, low sample sizes are likely to decrease the sensitivity and accuracy of DL models to predict metabolic phenotypes [14,55,80]. The limited number of labeled patient samples is a common issue for training any type of ML or DL model [81]. This, together with the significantly greater number of features describing each sample, negatively impacts unsupervised learning. An improved performance in the ML models while predicting individual patient responses to therapeutic drugs has been achieved by iteratively combining supervised models and feature selection [82,83]. In the context of DL models applied to image datasets, few-shot classification models are gaining attention owing to their architectural characteristics to learn from smaller dataset sizes [33,38]. Nonetheless, rapid advancements in high-throughput metabolomics, together with an efficient DL approach, might ameliorate these shortcomings.

A graph convolutional neural network (GCNN) is a type of deep neural network that works directly with a graph structure. A GCNN can be a DNN with fully connected convolutional layers. Node categorization is commonly used in the GCNN [23]. In essence, each node in the network has a label, and the task is to predict the labels of unknown nodes based on other node information. The GCNN model was developed as a more flexible way of propagating information through many tiers of neighboring nodes in a graph network. Traditional CNNs are inefficient in capturing complex neighborhood information because they assess small regions based on a fixed convolutional kernel, resulting in limited performance and interpretability of functional and structural feature analyses [22,23]. ML models that employ a graph network have an advantage in that they can accurately describe complex physical entities and processes, as well as irregular interactions. A GCNN was used to conduct pan-cancer analysis to detect the likelihood of tumor metastasis. Gene regulatory networks and gene expression features were extracted using a GCNN. The high-dimensional features of each sample were represented as image pixels and utilized to predict the likelihood of tumor metastasis in each patient [84].

Metastasis is the most common cause of cancer-related death. Monitoring variations in cancer gene expression is crucial for understanding metastatic pathways for tumors and cellular activities. A GCNN was recently used to combine the image features of metastasis to local lymph nodes for cancer trajectory prediction and clinical decision-making [21]. A GCNN can utilize the gene expression data of a metastatic and non-metastatic cohort as a graph signal molecular network and compute an individual subnetwork for a test patient cohort (Fig. 5). In summary, clinical metadata from the diseased and control groups (Fig. 5a) along with corresponding gene expression and/or other multi-omics data (Fig. 5b) as nodes, could be utilized to train a customized GCNN prediction model. This enabled the identification of patient-specific genes or feature subnetworks to deploy precision medicine (Fig. 5c-d). The differential expression of genes between training cohorts using cross-validation can optimize a GCNN model and identify the labels of network nodes, which correspond to potent biomarkers for metastasis (Fig. 5). Graph-CNN-based modeling presents an effective DL-based approach for selecting features, which take advantage of past information and generate patient-specific subnetworks, thereby influencing the outcome of individual categorization. For instance, a GCNN was employed to integrate patients' clinical-genomic data for pan-cancer analysis and investigate the stratification of lung tumors for immunotherapy [85]. The GCNN has the advantage of flexibility in clinical metadata integration, which may include clinical subtypes as well as patient-specific genes that might be associated with malignant or benign phenotypes. A variation of GCNN was utilized to predict the effect of pharmacotherapy on *in vitro* cancer cell growth [22]. Correspondingly, by using a GCNN, researchers have been able

to combine different features and extract cancer-related traits from gene expression data with significantly improved prediction accuracy [84]. Together, these results raise the possibility of using the GCNN as an effective DL method to examine the clinical effectiveness of cancer diagnostics and treatments.

5. Deep learning-based ensemble modeling for precision cancer medicine

One of the major aims of cancer precision medicine is to identify the molecular traits responsible for certain clinical outcomes. At present, a strategy involving molecular traits through drug-specific informative genes remains a popular means to investigate sensitivity or resistance to cancer drugs [81]. DL approaches have a myriad of modeling architectures, making it difficult to determine the best approach. Occasionally, it might be feasible to undertake a systematic study of all possible gene candidates using microarray gene expression data to identify new potential biomarkers. However, feature selection based only on statistical processing and single data types often does not account for personalized molecular information and clinical metadata, resulting in the selection of genes or traits with limited biological relevance [86]. Recurrent neural network and long-short-term memory (LTMS) models are good at analyzing repetitive patterns in sequences and time-series forecasting, whereas CNNs are better at detecting image features [87–89]. When omics data must be computed together with histological data and other clinical metadata, it is reasonable to adopt a DL-based multi-model ensemble network for biomarker discovery. A previous study suggested that ensemble-based modeling performs well in breast cancer diagnosis, in terms of both stability and prediction accuracy because of the collective representation of multi-model prediction schemes [4]. A recent study in breast cancer patients implementing bimodal DNN with heterogeneous gene expression data and clinical information outperformed all prognosis prediction models, as indicated by survival analysis and model performance, with the potential for the development of precision medicine [86]. Rather than depending on single-model feature prediction, an unsupervised network-based selector combined with DL models may enhance robustness. This is especially important in precision medicine, where a patient's clinical information might have a significant impact on one model, while having little or no effect on others. Fig. 6 summarizes how multi-omics data, along with clinical metadata, can be integrated into an ensemble-based multimodal network to identify potential biomarkers and facilitate precision medicine. Biomarker discovery involves an ensemble-based multi-omics feature selection strategy, and the efficacy of the predictor is evaluated on the basis of clinical survival analysis. This strategy is highly beneficial for precision medicine, as patients' clinical metadata can be directly correlated to the outcome of survival analysis; thus, the best performing models within the ensemble will have more influence on biomarker discovery and disease prediction. Taken together, these studies suggest that DL-based ensemble modeling can be advantageous when single data-type modeling is unable to achieve high performance. A limitation of multi-omics ensemble-based models is that they are computationally intensive and, therefore, difficult to deploy in clinics. Additionally, the generation of multi-omics data requires sophisticated instruments, and the acquisition of erroneous data may occur owing to a lack of expertise, ultimately reducing the performance of ensemble models.

6. Combination of deep learning and deep reinforcement learning

DL models learn from a training dataset and then apply that knowledge to infer classes in an entirely new dataset, whereas reinforcement learning (RL) models learn dynamically using

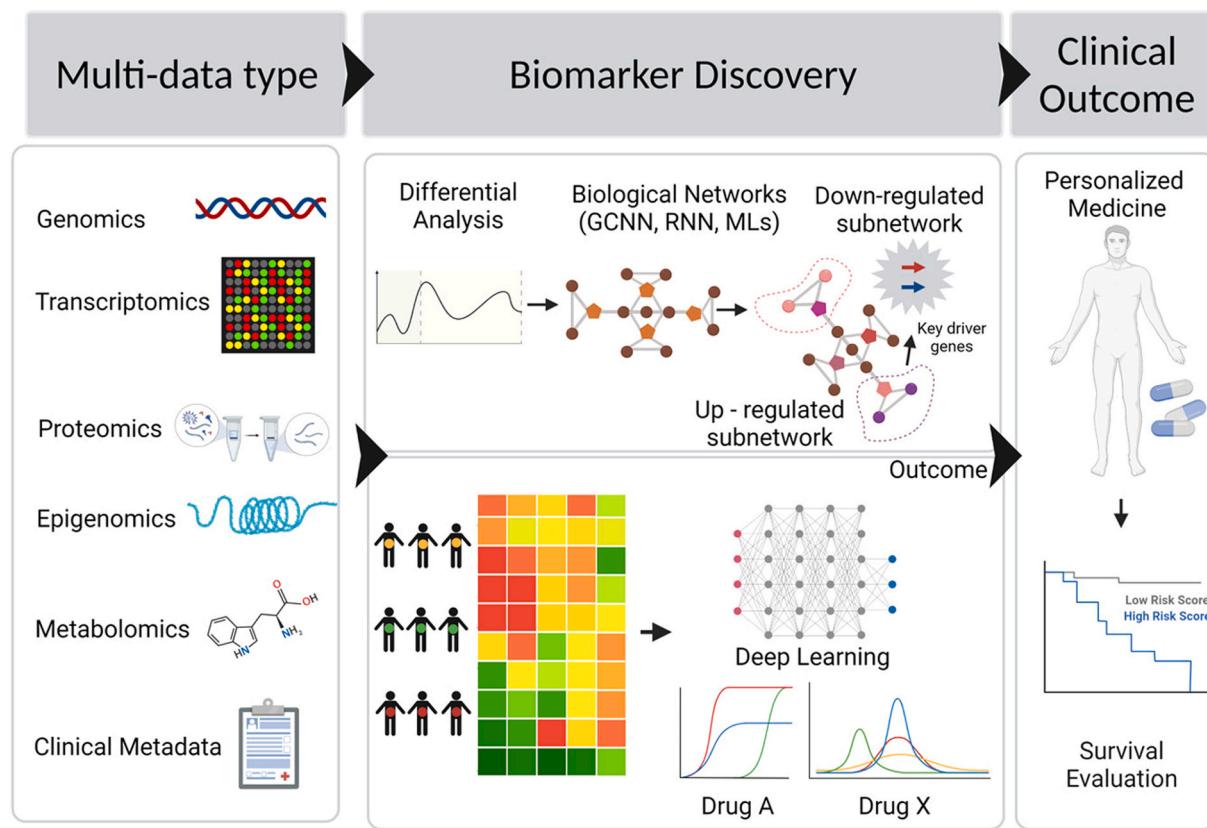


Fig. 6. Deep learning ensemble modeling for precision medication. Ensemble modeling can combine multiple data types along with clinical metadata with a variety of conventional ML and DL models. It predicts the final scores based on the ‘voting’ of multiple models, which makes the results resilient to single-point model failure. Patients’ survival outcomes and treatment efficacy output for an investigational medication can be fed back to the ensemble model for optimization of precision medicine. GCNN, graph convolution neural network; RNN, recurrent neural network; ML, machine learning.

continuous feedback to optimize a predesignated reward [90–92]. Deep reinforcement learning (DRL) combines the features of both RL and DL models to achieve optimum results in solving computationally intensive tasks, such as real-world object recognition, molecular structure simulation, and gaming. DRL has been used for applications ranging from computer gaming to medical technologies by incorporating DL models into reinforcement learning algorithms [93–95]. NVIDIA, a well-known gaming technology company, is actively involved in the development of AI systems for medical research, which illustrates the growing interest in DL and DRL in this field [96–98]. In terms of innate immunity, Petersen et al. used DRL to develop an adaptive precision medicine policy that defines effective multi-cytokine therapy for sepsis patients [99]. Cancer-related omics datasets using DRL have been able to address a wide range of sequential decision-making problems, suggesting their potential in cancer research [92,96]. DRL was utilized on existing therapeutic methods to build automated radiation adaptation procedures for patients with non-small cell lung cancer (NSCLC). It aims to maximize localized tumor growth control while reducing radiation dose and inhibiting radiation-induced pneumonitis [94]. Furthermore, researchers have explored DRL’s potential for correlating distinct forms of cancer associated with tumor protein 53 (TP53) mutation patterns, and their unique impacts on tumors [100]. Recent developments in advanced quantum DRL have provided a framework for clinical decision-making that can estimate personalized radiotherapy and propose the best dosage adjustment strategy [101]. The implementation of DRL and DL enables the design of sophisticated disease models that takes input from real-world complex data and are not limited to certain pre-described protocols. These trends indicate a rapid adoption of DRL and DL to enhance the potential of AI in cancer research.

7. Summary and Outlook

Clinical research is increasingly demonstrating the broad applicability of DL methods, including in cancer diagnosis and therapy. Recent studies have demonstrated the potential of DL-based ensemble models to augment personalized patient treatment strategies based on medical histories and diagnostics. Advancements in high-throughput and large-scale omics assays have generated large volumes of data that can be utilized for biomarker discovery. Furthermore, precision medicine-based therapeutic interventions for cancer have shown promising efficacy when using a DL-based multimodal ensemble strategy that involves gene-protein-metabolite interaction (Fig. 7). These approaches involve *de novo* drug synthesis, candidate gene clustering, and clinical outcome prediction based on a set of biomarkers discovered by processing complex features from multi-omics datasets.

However, DL models have several limitations, including small sample sizes, lack of interpretability, reliability of computational resources, and scarcity of professional expertise [102]. Training a DL model requires large numbers of well-characterized samples, and a small sample size has been shown to decrease the predictability and accuracy of DL models. Furthermore, the quality of the training data is critical because down- or up-sampling of an imbalanced dataset may introduce biases into model predictions. Complex DL models that are applied to large volumes of data, including imaging datasets, must be trained iteratively, which is computationally expensive. In this context, consensus neural networks can provide effective interrogation of data and may aid in reducing features. In other cases, conventional ML methods can be trained faster than DL models [103]. DL-based prediction requires manual curation; thus, the selection and validation of molecular features and signatures using a

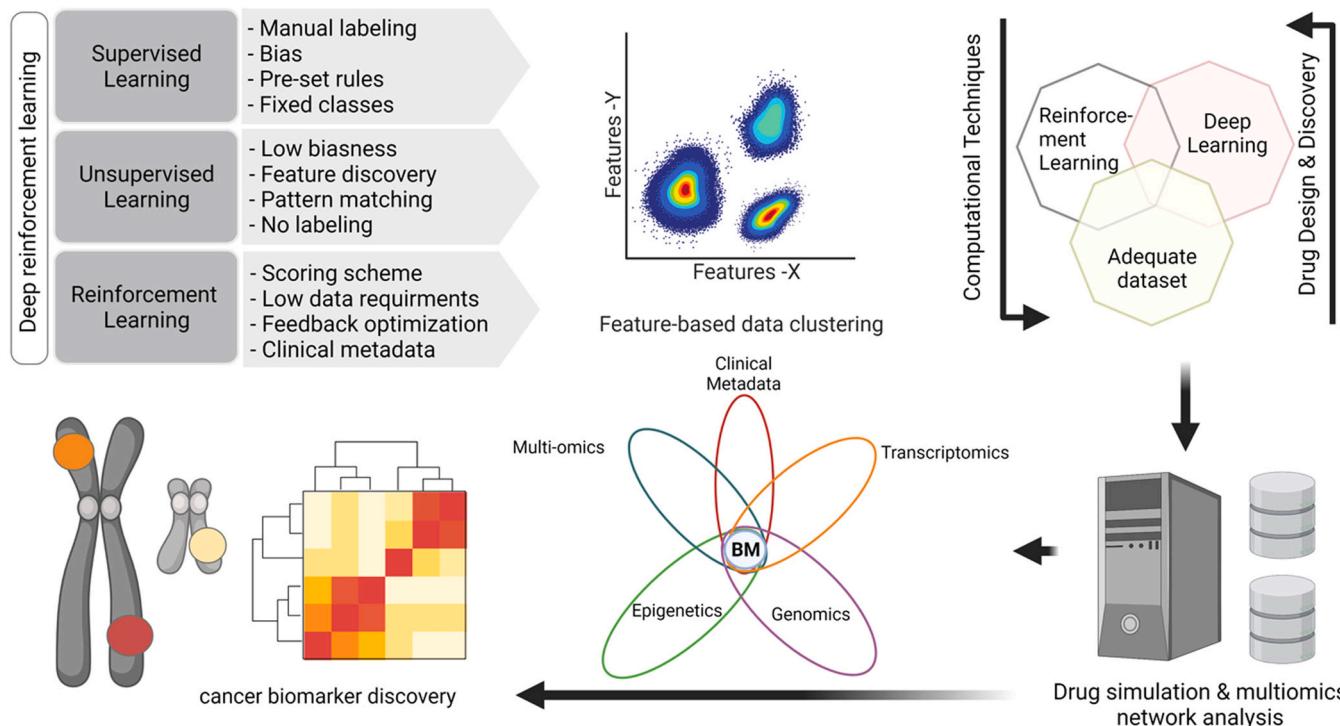


Fig. 7. Conventional modeling and deep learning framework for precision medicine and drug discovery. Knowledge from DL and DRL techniques in the presence of adequate datasets can assist in cancer biomarker discovery. The combination of DL and DRL methods shows promising potential to facilitate the development of anti-cancer drugs based on disease profiles to assist precision medication using multi-omics datasets and clinical metadata. BM, biomarker; DL, deep learning; DRL, deep reinforcement learning.

DL model remain pertinent to biomarker discovery in cancer research. In conclusion, we anticipate that DL and AI have great potential for identifying biomarkers for cancer and in other medical areas as well.

CRediT authorship contribution statement

Vivek Bhakta Mathema: Conceptualization, Writing – original draft. **Partho Sen:** Conceptualization, Writing – original draft. **Santosh Lamichhane:** Writing – original draft. **Matej Orešić:** Writing – original draft. Sakda Khoomrung: Conceptualization, Writing – original draft.

Declarations of interest

None.

Acknowledgements

SK was supported by a Chalermphrakiat Grant from the Faculty of Medicine, Siriraj Hospital. SK and VBM were supported by the Mahidol University Grant No. (IO) R016420001 (S.K.). This research received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research, and Innovation [grant number B16F640099]. We thank Dr. Yotsawat Pomyen from the Chulabhorn Research Institute, Bangkok, Thailand, and Dr. Jonathan L. Robison from the BioInnovation Institute, Denmark, for their comments on the manuscript.

References

- [1] Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS* 2010;5:463–6.
- [2] Broza YY, Zhou X, Yuan M, et al. Disease detection with molecular biomarkers: from chemistry of body fluids to nature-inspired chemical sensors. *2019;119:11761–817.*
- [3] Martinkova J, Gadher SJ, Hajduch M, et al. Challenges in cancer research and multifaceted approaches for cancer biomarker quest. *2009;583:1772–84.*
- [4] Hsieh S-L, Hsieh S-H, Cheng P-H, et al. Design ensemble machine learning model for breast cancer diagnosis. *2012;36:2841–7.*
- [5] Poirion OB, Jing Z, Chaudhary K, et al. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med* 2021;13:112.
- [6] Badré A, Zhang L, Muchero W, et al. Deep neural network improves the estimation of polygenic risk scores for breast cancer. *2021;66:359–69.*
- [7] Sud A, Turnbull C, Houlston RJNpo. Will polygenic risk scores Cancer ever be Clin Useful? *2021;5:1–5.*
- [8] Chen H, Dou Q, Yu L, et al. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images 2018;170:446–455.
- [9] Inglese P, McKenzie JS, Mroz A, et al. Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. *Chem Sci* 2017;8: 3500–11.
- [10] Tornig W, Altman RB, J. 3D deep convolutional neural networks for amino acid environment similarity analysis 2017;18:1–23.
- [11] Cui P, Li X, Huang C, et al. Metabolomics and its Applications in Cancer Cachexia 2022;9.
- [12] Chen D, Liu S, Kingsbury P, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data 2019;2:1–5.
- [13] Bomanne A, Gonçalves A, Ballester P.J.J.Fig. Paclitaxel response can be predicted with interpretable multi-variate classifiers exploiting DNA-methylation and miRNA data 2019;10:1041.
- [14] Grapov D, Fahrmann J, Wanichthanarak K, et al. Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *OMICS* 2018;22:630–6.
- [15] Alakwaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res* 2018;17:337–47.
- [16] L. Zhang C, Lv Y, Jin et al. Deep Learn-Based multi-omics data Integr Reveals two Progn subtypes High-risk neuroblastoma 2018 477.
- [17] Malik V, Kalakoti Y, Sundar D.J.Bg Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer 2021;22:1–11.
- [18] Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. *Seminars in Nuclear Medicine*. Elsevier; 2022.
- [19] Kang M, Ko E, Mersha T.B.J.Bi.B. A roadmap for multi-omics data integration using deep learning 2022;23:bbab454.
- [20] Melnikov A.D., Tsentalovich Y.P., Yanshole V.V.J.Ac Deep learning for the precise peak detection in high-resolution LC–MS data 2019;92:588–592.
- [21] Kazmierski M, Haibe-Kains B.Japa. Lymph Node Graph Neural Networks for Cancer Metastasis Prediction 2021.
- [22] Singha M, Pu L, Busch K, et al. GraphGR: A graph neural network to predict the effect of pharmacotherapy on the cancer cell growth 2020.

- [23] Ahmedt-Aristizabal D, Armin MA, Denman S et al. Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future 2021.
- [24] Prykhodko O, Johansson S.V., Kotsias P.-C. et al. A de novo molecular generation method using latent vector based generative adversarial network 2019;11:1–13.
- [25] Méndez-Lucio O., Baillif B., Clevert D.-A. et al. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence 2020;11:1–10.
- [26] Hassan M., Ali S., Alquhayz H. et al. Developing intelligent medical image modality classification system using deep transfer learning and LDA 2020;10:1–14.
- [27] Puttagunta M., Ravi SJMT, Applications. Medical image analysis based on deep learning approach 2021;80:24365–24398.
- [28] deSouza N.M., Achten E., Alberich-Bayarri A. et al. Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL® subcommittee of the European Society of Radiology (ESR) 2019;10:1–16.
- [29] Pataki BA, Olar A, Ribli D, et al. HunCRC: annotated pathological slides to enhance deep learning applications in colorectal cancer screening. 2022;9:370.
- [30] Alzubaidi L, Al-Shamma O., Fadhel M.A. et al. Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model 2020;9:445.
- [31] Lakshmanaprabu S., Mohanty S.N., Shankar K. et al. Optimal deep learning model for classification of lung cancer on CT images 2019;92:374–382.
- [32] Gulum M.A., Trombley C.M., Kantardzic M.J.A.S. A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging 2021;11:4573.
- [33] Feyjje A.R., Azad R., Pedersoli M. et al. Semi-supervised few-shot learning for medical image segmentation 2020.
- [34] Wen L, Li X, Li X, et al. A new transfer learning based on VGG-19 network for fault diagnosis. 2019 IEEE 23rd Int Conf Comput Support Coop Work Des (CSCWD) 2019:205–9.
- [35] Gao Y., Cui YJNc. Deep transfer learning for reducing health care disparities arising from biomedical data inequality 2020;11:1–8.
- [36] Ho N., Kim Y-CJSr. Evaluation of transfer learning in deep convolutional neural network models for cardiac short axis slice classification 2021;11:1–11.
- [37] Wu Z., Wang L., Li C. et al. DeepLRHE: A Deep Convolutional Neural Network Framework to Evaluate the Risk of Lung Cancer Recurrence and Metastasis From Histopathology Images 2020;11.
- [38] Teng H., Zhang W., Wei J. et al. Few-shot Learning on the Diagnosis of Lymphatic Metastasis of Lung Carcinoma 2021.
- [39] Shim WS, Yim K, Kim T-J, et al. DeepRePath: identifying the prognostic features of early-stage lung adenocarcinoma using multi-scale pathology images and deep convolutional neural networks. 2021;13:3308.
- [40] Wang Y., Yao Q., Kwok J.T. et al. Generalizing from a few examples: A survey on few-shot learning 2020;53:1–34.
- [41] Sharma A., Vans E., Shigemizu D. et al. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture 2019;9:1–7.
- [42] Mostavi M., Chiu Y.-C., Huang Y. et al. Convolutional neural network models for cancer type prediction based on gene expression 2020;13:1–13.
- [43] Saldanha O.L., Quirke P., West N.P. et al. Swarm learning for decentralized artificial intelligence in cancer histopathology 2022;1:8.
- [44] Liu S., Masurkar A.V., Rusinek H. et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs 2022;12:17106.
- [45] Kim J.S., Han J.W., Bae J.B. et al. Deep learning-based diagnosis of Alzheimer's disease using brain magnetic resonance images: an empirical study 2022;12:18007.
- [46] Arshad Q.A., Ali M., Hassan S.-u et al. A dataset and benchmark for malaria life-cycle classification in thin blood smear images 2022;1–13.
- [47] Wu X., Wang H.-Y., Shi P. et al. Long short-term memory model—a deep learning approach for medical data with irregularity in cancer predication with tumor markers 2022;144:105362.
- [48] Lee TY, Huang KY, Chuang CH, et al. Incorporating deep learning and multi-omics autoencoding for analysis of lung adenocarcinoma prognostication. *Comput Biol Chem* 2020;87:107277.
- [49] Zhang X., Wang J., Lu J. et al. Robust prognostic subtyping of muscle-invasive bladder cancer revealed by deep learning-based multi-omics data integration. *Front Oncol* 2021;11:689626.
- [50] Arjmand B., Hamidpour S.K., Tayanloo-Beik A. et al. Machine Learning: A New Prospect in Multi-Omics Data Analysis of Cancer. *Frontiers in Genetics* 2022;13.
- [51] Petrovsky DV, Kopylov AT, Rudnev VR, et al. Managing of unassigned mass spectrometric data by neural network for cancer phenotypes classification. *J Pers Med* 2021;11.
- [52] Hira MT, Razzaque MA, Angione C, et al. Integrated multi-omics analysis of ovarian cancer using variational autoencoders. *Sci Rep* 2021;11:6265.
- [53] Chaudhary K, Poirion OB, Lu L, et al. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 2018;24:1248–59.
- [54] Wang T., Shao W., Huang Z. et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification 2021;12:1–13.
- [55] Pomyen Y, Wanichthanarak K, Poungsombat P, et al. Deep metabolome: applications of deep learning in metabolomics. *Comput Struct Biotechnol J* 2020;18:2818–25.
- [56] Lin E, Lin C-H, Lane, YJM H-. Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. 2020;25:3250.
- [57] Kopylov AT, Petrovsky DV, Stepanov AA, et al. Convolutional neural network in proteomics and metabolomics for determination of comorbidity between cancer and schizophrenia. *J Biomed Inf* 2021;122:103890.
- [58] Tran K.A., Kondrashova O., Bradley A. et al. Deep learning in cancer diagnosis, prognosis and treatment selection 2021;13:1–17.
- [59] Kleppe A, Skrede O-J, De Raedt S, et al. Designing deep learning studies in cancer diagnostics. 2021;21:199–211.
- [60] Pomyen Y, Wanichthanarak K, Poungsombat P, et al. Deep metabolome: applications of deep learning in metabolomics. 2020;18:2818–25.
- [61] Troisi J, Colucci A, Cavallo P, et al. A serum metabolomic signature for the detection and grading of bladder. *Cancer* 2021;11:2835.
- [62] Hassan MA, Al-Sakkaf K, Shait Mohammed MR, et al. Integration of transcriptome and metabolome provides unique insights to pathways associated with obese breast cancer patients. 2020;10:804.
- [63] O'Boyle N., Dalke A. DeepSMILES: an adaptation of SMILES for use in machine-learning of chemical structures 2018.
- [64] Kimber T.B., Gagnebin M., Volkamer AJAlitLS. Maxsmi: maximizing molecular property prediction performance with confidence estimation using SMILES augmentation and deep learning 2021;1:100014.
- [65] Mouchlis VD, Afantitis A, Serra A, et al. Advances in de novo drug design: from conventional to machine learning methods. 2021;22:1676.
- [66] Hert J, Irwin JJ, Laggner C, et al. Quantifying biogenetic bias in screening libraries. 2009;5:479–83.
- [67] El Rassy E, Delaroque A, Sambou P, et al. On the potential of the particle swarm algorithm for the optimization of detailed kinetic mechanisms. comparison with the genetic algorithm. 2021;125:5180–9.
- [68] Goodfellow I., Pouget-Abadie J., Mirza M. et al. Generative adversarial nets 2014;27.
- [69] Sousa T, Correia J, Pereira V, et al. Generative deep learning for targeted compound design. 2021;61:5343–61.
- [70] Shayakhmetov R, Kuznetsov M, Zhebrak A, et al. Molecular generation for desired transcriptome changes with adversarial autoencoders. 2020;11:269.
- [71] Janiesch C., Zschech P., Heinrich K.J.E.M. Machine learning and deep learning 2021;31:685–695.
- [72] Liang G, Fan W, Luo H, et al. The emerging roles of artificial intelligence in cancer drug development and precision therapy. 2020;128:110255.
- [73] Chang Y., Park H., Yang H.-J. et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature 2018;8:1–11.
- [74] Pan X., Lin X., Cao D. et al. Deep learning for drug repurposing: Methods, databases, and applications 2022:e1597.
- [75] Issa NT, Stathias V, Schürer S, et al. Machine and deep learning approaches for cancer drug repurposing. *Seminars in cancer biology*. Elsevier.; 2021. p. 132–42.
- [76] Zhou Q., Yong B., Lv Q, et al. Deep autoencoder for mass spectrometry feature learning and cancer detection. *IEEE Access* 2020;8:45156–66.
- [77] Shaffie A, Soliman A, Fu XA, et al. A novel technology to integrate imaging and clinical markers for non-invasive diagnosis of lung cancer. *Sci Rep* 2021;11:4597.
- [78] Fang X., Liu Y., Ren Z, et al. Lilikoi V2.0: a deep learning-enabled, personalized pathway-based R package for diagnosis and prognosis predictions using metabolomics data. *Gigascience* 2021;10.
- [79] Al-Akwaa FM, Yunits B, Huang S, et al. Lilikoi: an R package for personalized pathway-based classification modeling using metabolomics data. *Gigascience* 2018;7.
- [80] Sen P, Lamichhane S, Mathema VB, et al. Deep learning meets metabolomics: a methodological perspective. *Brief Bioinform* 2021;22:1531–42.
- [81] Parca L., Pepe G., Pietrosanto M. et al. Modeling cancer drug response through drug-specific informative genes 2019;9:1–11.
- [82] Naulaerts S., Menden M.P., Ballester P.J.J.B. Concise polygenic models for cancer-specific identification of drug-sensitive tumors from their multi-omics profiles 2020;10:963.
- [83] Huang C., Clayton E.A., Matyunina L.V. et al. Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy 2018;8:1–8.
- [84] Xu Y., Cui X., Wang YJFic et al. Pan-cancer metastasis prediction based on Graph deep learning method 2021;9:1133.
- [85] Fang C., Xu D., Su J, et al. DeepPan: deep patient graph convolutional network integrating clinicogenomic evidence to stratify lung cancers for immunotherapy 2021;4:1–10.
- [86] Cheng L.-H., Hsu T.-C., Lin C.J.S.R. Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction 2021;11:1–10.
- [87] Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions 2021;8:1–74.
- [88] Wang T., Tian Y., Qiu RGJjob et al. Long short-term memory recurrent neural networks for multiple diseases risk prediction by leveraging longitudinal medical records 2019;24:2337–2346.
- [89] Petneházi Gjapa. Recurrent neural networks for time series forecasting 2019.

- [90] Wang M, Zhang Q, Lam S, et al. A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning 2020;10:2177.
- [91] Chen J,X,J,Ci,S., Engineering. The evolution of computing: AlphaGo 2016;18:4–7.
- [92] François-Lavet V, Henderson P, Islam R, et al. An introduction to deep reinforcement learning 2018.
- [93] Moreau G, François-Lavet V, Desbordes P, et al. Reinforcement learning for radiotherapy dose fractioning automation. 2021;9:214.
- [94] Liu Z, Yao C, Yu H, et al. Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things. Future Gener Comput Syst 2019;97:1–9.
- [95] Jonsson AJKD. Deep reinforcement learning in medicine. arXiv:2103.05115 2019;5:18–22.
- [96] Iqbal MJ, Javed Z, Sadia H, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. Cancer Cell Int 2021;21:1–11.
- [97] Pandey M, Fernandez M, Gentile F, et al. The transformational role of GPU computing and deep learning in drug discovery. Nat. Mach. Intel. 2022;4:211–21.
- [98] Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. International MICCAI Brainlesion Workshop. Springer.; 2018. p. 311–20.
- [99] Day JD, Cockrell C, Namas R, et al. Inflammation and disease: modelling and modulation of the inflammatory response to alleviate critical illness. 2018;12:22–9.
- [100] Rahmani A, Minaei-Bidgoli B, Ahangaran M. Deep reinforcement learning for diagnosing various types of cancer by TP53 mutation patterns. Res. Sq. 2021.
- [101] Niraula D, Jamaluddin J, Matuszak MM, et al. Quantum deep reinforcement learning for clinical decision support in oncology. Appl Adapt Radiother 2021;11:1–13.
- [102] Grapov D, Fahrmann J, Wanichthanarak K, et al. Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. OMICS: A J Integr Biol 2018;22:630–6.
- [103] Chen D, Liu S, Kingsbury P, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. NPJ Digit Med 2019;2:43.

Glossary

- AI:** artificial intelligence
ANN: artificial neural network
AUROC: receiver operating characteristic curve
CDSS: clinical decision support systems
CNN: convolutional neural networks
CT: computed tomography
DESI: Desorption electrospray ionization
DL: deep learning
DNN: deep neural networks
DRL: Deep reinforcement learning
GAN: generative adversarial network
GCNN: graph convolutional neural network
GNN: graph neural networks
HCC: hepatocellular carcinoma
KRT7: Keratin 7
LTMS: long-short-term memory
LUAD: lung adenocarcinoma
MIBC: muscle-invasive bladder cancer
ML: machine learning
MRI: magnetic resonance imaging
MS: mass spectrometry
NSCLC: non-small cell lung cancer
RL: reinforcement learning
RNN: recurrent neural networks
RTART: recursive partitioning and regression trees
SL: swarm learning
SMILES: simplified molecular-input line-entry system
TCGA: the cancer genome atlas
TL: transfer learning
TP53: tumor protein 53