Review

Cancer genetics and deep learning applications for diagnosis, prognosis, and categorization

Massoud Sokouti[1,2,3], Babak Sokouti[4]*

[1]Research Center of Evidence-Based Medicine, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Health Promotion Research Center, Tabriz Medical Sciences, Islamic Azad University, Tabriz, Iran
[3]Department of Physiology, Faculty of Medicine, Tabriz Medical Sciences, Islamic Azad University, Tabriz, Iran
[4]Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract

Gene expression data are used to discover meaningful hidden information in gene datasets. Cancer and other disorders may be diagnosed based on differences in gene expression profiles, and this information can be gleaned by gene sequencing. Thanks to the tremendous power of artificial intelligence (AI), healthcare has become a significant user of deep learning (DL) for predicting cancer diseases and categorizing gene expression. Gene expression Microarrays have been proved effective in predicting cancer diseases and categorizing gene expression. Gene expression datasets contain only limited samples, but the features of cancer are diverse and complex. To overcome their dimensionality, gene expression datasets must be enhanced. By learning and analyzing features of input data, it is possible to extract features, as multidimensional arrays, from the data. Synthetic samples are needed to strengthen the range of information. DL strategies may be used when gene expression data are used to diagnose and classify cancer diseases.

Keywords: Gene expression, High dimensionality, Deep learning, Cancer

1. INTRODUCTION

Characterized by uncontrolled, abnormal cell proliferation and spread to other tissues and organs, cancer is a heterogeneous disease caused by random somatic mutations. Classification of cancer subtypes based on their characteristics is conducive to cancer research. Current classification methods commonly rely on gene expression data [1]. Prior studies showed that identifying and addressing cancer at its initial stages can avert about 30 – 50% of cancer-related fatalities [2]. In recent years, precision medicine has made significant progress, in general, and in oncology, in particular. Researchers have evaluated drug efficacy using established cancer cell lines in laboratory settings. They now incorporate genomic data, gene expression analysis, and cheminformatics to predict how patients may respond to specific treatments. Considering each patient's distinct genetic and molecular profiles, it is of great importance to personalize medical therapies for each patient [3].

Thanks to the development of computer-aided technology, machine-learning techniques have substantially improved detection and prediction accuracy [4]. For instance, molecular biomarkers can help oncologists determine the treatment intensity of an individual patient. Based on the disease profile of a particular patient group, predictive biomarkers can be used to adjust the intensity of therapies. Moreover, predictive biomarkers can help clinicians select the optimal treatment. In contrast, routine managements cost more, last longer, and require more tissues since numerous related molecular biomarkers have to be determined, such as in the cases of non-small-cell lung cancer [5].

Gene expression data from tumor tissue samples can be used to determine if molecular factors contribute to the progression of a disease or influence a patient's survival. Efficient data extraction allows for control over

***Corresponding author:**
Babak Sokouti (sokoutib@tbzmed.ac.ir)

Received: 17 June 2024; Revision received: 10 July 2024; Accepted: 22 July 2024; Published: 09 August 2024

How to cite this article: Sokouti M, Sokouti B. Cancer genetics and deep learning applications for diagnosis, prognosis, and categorization. *J Biol Methods.* 2024;11(3): e99010017. DOI: 10.14440/jbm.2024.0016

diagnostic mechanisms, improving prognosis and increased significance [6]. It is estimated that nearly ten million people die of cancer each year. As a result, clinical practitioners, patients, researchers, and policymakers all attach great importance to survival prediction. Still, such prediction tends to be highly subjective and is subject to clinicians' intuition; therefore, it can only achieve limited accuracy [7].

Bioinformatics addresses health-care-related problems by applying computational techniques to process biological data. The expression of genes within a human cell can be determined by using RNA sequencing (RNA-seq) or Microarray/DNA Chip. Many samples from various types of cancers have been analyzed, and the corresponding data have been shared with the public. The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium database are two databases. Public databases for these kinds of studies include Stanford Microarray, GenBank, Array express, gene expression omnibus (GEO), and the National Center for Biotechnology Information [8].

Aside from studying the entire human genome, next-generation sequencing (NGS) can identify various genomic alterations. With these techniques, several limitations associated with microarray experiments for determining gene expression have been overcome [9]. By utilizing NGS techniques [10], it is possible to detect both coding and non-coding RNA to study the entire human genome. Advances in structural genomics have made it possible to study the whole human genome. DNA methylation is a promising biomarker for detecting and classifying cancer. It involves some activities, including imprinting of genomic DNA, inactivating X-chromosomes, and suppressing transcription and transposition of repetitive elements [11]. Although total RNA sequencing can be used for cancer diagnosis, the complexity of the data involved makes it a tremendous challenge for traditional machine-learning techniques to evaluate genetic variants. For this reason, a high-dimensional dataset requires more data for analysis each time. Information is maintained when various variables (or features) are projected onto a smaller number, but processing takes weeks [8].

One can determine the relationships between biological molecules in an organism and identify the different types of cancer by analyzing Omics data. Because of the relatively small amount of available samples and the high number of features, genomic data are difficult to interpret. The accuracy of gene expression prediction in clinical applications remains to be improved. Several studies have suggested that gene expression may help predict breast cancer recurrence and metastasis. Redundant, noisy, and irrelevant data are expected in gene expression values, including non-reproducible, overfitting, noisy, and non-reproducible results. Two approaches, namely feature selection (FS) and feature extraction, address these issues.

Semi-supervised, supervised, and unsupervised methods are among the most popular informative FS types for retrieving biological data. FS strategies often combine ensemble-, filter-, embedded-, and hybrid-based models [12].

Since gene expression data have grown exponentially, several approaches for analyzing and diagnosing cancer using machine learning have been introduced. These methods use gene expression data to classify samples according to their predicted survival status by analyzing their gene expression data. Classical machine learning methods, such as the Cox proportional hazard model and the support vector machine method have been widely applied to predict and diagnose cancer [13].

Due to its high-dimensional, heterogeneous, and complex nature, deep learning (DL) can be helpful in the discovery of biomarkers, categorization of patients, genomic analysis, and sequencing. A suitable kernel for current methods has not yet been established [14]. In general, gene expression datasets are subject to problems, such as small datasets, high dimensionality, and unbalanced data, and identifying significant genes can be addressed by gene selection. Furthermore, high dimensionality or multidimensional arrays refer to data structures that are beyond the typical one or more than two dimensions. Difference analysis is the most common method for determining gene significance, but it can be challenging when the data set is small and of high dimension [15]. A DL -based approach is proposed instead of a statistics-based approach with a "hard" cutoff to reduce feature information loss. Image processing, voice recognition, natural language processing, and chemical pattern recognition are among the applications of DL that have been developed in recent years. Cancer prognostication has also been predicted to use DL, but prognostic genes have been less studied for this purpose [16].

2. GENETICS OF CANCER

Inheritance or acquisition of at least one genetic alteration causes cancer cells to multiply uncontrollably. Gene-encoding proteins associated with cell cycle or DNA recovery systems may be affected when cancer-causing gene changes interfere with the replication and life-cycle of normal cells. Whenever oncogenic signals are corrupted, genomic instability results, and tumor cells are more likely to develop genetic mutations, and further cause tumor growth and spread [17].

A genetic mutation causes an uncontrollable growth of cells and leads to cancer. This mutation disrupts genes that control cell division and DNA repair, thereby resulting in genomic instability, facilitating tumor growth, and metastasis by establishing oncogenic signaling pathways.

Cancer cells can increase in number when one or more genetic changes occur. Lifespan is shortened due to changes

in deoxyribonucleic acid production when genes associated with cell proliferation are eliminated. Cancerous traits can be induced by altering its signaling pathway, leading to genomic instability [17].

Cancer-related genes and non-cancer-related genes are two types of somatic genes contributing to oncogenesis and tumor progression. Proto-oncogenes are characterized by their ability to generate novel proteins that, through biochemical cascades and bio-cellular processes, control cell division, prevent differentiation, or delay apoptosis. In addition, increased expression of a proto-oncogene leads to the formation of an oncogene, which can be triggered by changes or magnification of a gene (multiple copies) (e.g., human epidermal growth factor receptor 2 and RAS). A tumor suppressor gene or gatekeeper gene is another gene that is commonly altered in cancer cells' genomes (i.e., BRCA1, BRCA2, p53, or TP53) [18]. On the other hand, proto-oncogenes transform when their functions are altered [19]. In contrast, tumor suppressor genes handle oncogenesis when their wild-type copies are inactivated or lose function.

Aberrant gene silencing and activation because of epigenetic defects is a simplified model of genomic instability [20]. Hassanpour and Dehghani identified epigenetic changes to the DNA structure without altering its sequence directly. This study found mutations in transcription factors and non-coding RNAs, which regulate gene expression and repression and modulate various biological processes [21]. The DNA writer or the DNA eraser is a protein that recognizes DNA or histones altered by transferring or removing chemical groups. The reader is a protein that recognizes DNA or histones that have been changed. Thus, there is evidence for epigenetic diversity at the cellular level, since tumor cells display a spectrum of histone modifications both throughout the genome and within specific genes. Therefore, tumor genesis may result from a combination of epigenetic events [22].

New ways of diagnosing, categorizing, and treating cancer have emerged as a direct result of the advances in sequencing and understanding human genes. To evaluate the efficacy of targeted therapies, such as receptor tyrosine kinase inhibitors, knowledge about the disorder processes and the role of genetic variations in oncogenesis is essential. Cancer cells can experience a significant change from their original forms when a line of cells spreads and multiplies. As a cancer cell replicates, its genome will continue to change, and since it differs from its original genome, genetic mutations will accumulate. For a tumor to become heterogeneous, the cells in the primary cancer and the secondary multicellular clusters acquire new accumulated genetic modifications and become different in terms of their genetic composition as the tumor grows. Using tumor molecular profiling, specific genetic changes throughout the disease course can be identified to screen out effective targeted agents and detect drug resistance of tumors. Apoptosis impedes proliferation, and truncated differentiation and truncation are challenging to determine. By combining or modifying several genes, a single aberrant cell may develop into a full-blown, aggressive, invasive cancer [23].

Revolutionary new technologies are expected to lead the way in new modes for the prediction of genetic risk for cancer. Novel cancer susceptibility genes have been developed to help clinicians and researchers to "diagnose" previously undetected familial cancers through high-throughput screening. Turnbull *et al*., by work using high-throughput screening, demonstrated that the genomic organization of those prevalent tumors was much more complicated than was previously thought [24].

A system biologist can analyze biological events by considering a network-based system of interconnected parts. Such parts may include many diverse molecular and environmental elements that interact with each other on a variety of levels. Some examples of genetic variants impacting tumor activity include expression levels, protein motifs, and cellular networks. DNA and protein interaction failures in cancer cells' replication, transcription, metabolism, and signaling networks are crucial [18,25]. A schematic figure of cancer genetics has been illustrated in Figure 1.

3. TECHNOLOGICAL ADVANCES BASED ON GENOME

3.1. The sequencing of RNA

Recently, various approaches to RNA sequencing, including short-read, long-read, and direct RNA sequencing, as well as long-read complementary DNAs (cDNAs) and direct RNAs, were detailed. A paper expanded on short-read and long-read sequencing and direct RNA sequencing [26].

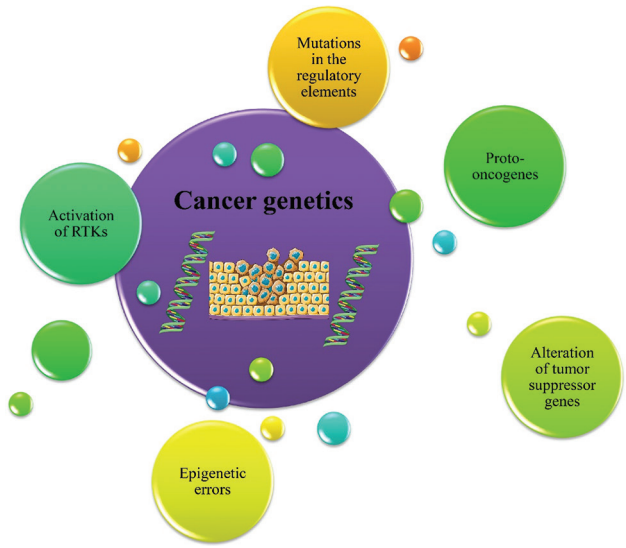

Figure 1. A schematic illustration of cancer genetics

Due to its higher quality data through the transcriptome and its lower cost, short-read cDNA sequencing is more affordable and more accessible for the application in the expression of RNA genes than Microarrays. In identifying and quantifying different isoforms, errors may appear during sample preparation and computational analysis, leading to false results. It is important to remember that the accuracy of genome mapping depends on the assembly of short RNA fragments as part of short-read sequencing. Using long-read sequencing, one can map the genome of mammalian cells whose transcripts are at least 1 – 2 kb but may exceed 100 kb since the sequencing software can identify large amounts of RNA and process it in its entirety. Errors caused by cDNA amplification and RNA-RNA chimeras may be avoided with long-read RNA sequencing (also known as dRNA-seq) since it does not involve the synthesis of cDNA. ONT developed this technique for this purpose [26, 27] and other nano pore sequencing techniques.

3.2. Gene expression arrays

Several genes expressed in cancer were first identified using microarrays created from spotted DNA amplifications from cDNA clones. Whole genome screens were developed when cDNA clones with sequence verification became available [28]. Identifying genes whose expression levels were influenced by subtle changes in phenotypes with the commercial arrays was difficult. Developing an oligonucleotide platform was motivated by the difficulties associated with the production and interpretation of cDNA arrays, which provide valuable information about genetic changes in cancer. The company developed unique sequences using oligonucleotide-based expression arrays through photolithography representing human genes on quartz wafers. Proprietary software with built-in smoothing and normalizing controls was used to calculate hybridization intensity. Over 55,000 probe sets represent over 20,000 genes on the U133Plus2 arrays. The complexity of these arrays has increased since the first 6800 transcripts were fully characterized, thanks to the higher quality of transcript databases. Quality control indicators can assess sample preparation. Repeat hybridization can reduce false discovery rates. Oligonucleotide array data processing has advanced dramatically. It is now possible to evaluate gene expression levels and investigate cancer development in various ways through oligonucleotide arrays [29]. These arrays are increasingly accessible and commercially available. One can determine how genes are expressed and deleted in cancer cells as part of an array. Much cancer research uses gene expression profiles, including comparing gene expression profiles across tumors at different stages of progression and comparing tumors with normal tissues to determine if gene expression changes [29,30].

3.3. An overview of data sources

RNA sequencing data and correlated clinical data of breast cancer patients can be downloaded from TCGA website (https://cancergenome.nih.gov/). From Xena Functional Genomics Explorer (https://xenabrowser.net/) clinical data from patients in the TCGA database can be downloaded, including progression-free survival and overall survival (OS) [31]. A transcriptome profile of a carcinoma sample and a standard sample could be obtained using TCGA. The GDC Data Transfer Tool (https://gdc.cancer.gov/access-data/gdc-data-transfertool) allowed for the generation and download of lncRNA sequencing data and clinical information [32].

A search of the largest libraries of gene expression data (e.g., GEO, ArrayExpress [33], and Oncomine [34] may help identify data generated using high-throughput methods (e.g., Microarrays, RNA-Sequencing). Finding those datasets within these repositories matched by specific search criteria is also possible. The data from ArrayExpress overlapped with those of GEO, which had already been identified, can also be retrieved [35] because ArrayExpress maintained a copy of all GEO datasets. Creating an account with Oncomine allows researchers to conduct dataset searches for free. Furthermore, included in the selected datasets from GEO are gene transcriptomes as well as genome-wide DNA methylation expression data in the tumor tissues and normal controls.

As a bonus, there is an application, Geo Cancer Prognostic Datasets Retriever, which is available for free on CPAN and GitHub and can be accessed through CPAN or GitHub for retrieving multiple gene expression datasets that predict the outcome of a specific cancer type from the GEO. Alameer and Chicco laid the foundations for several bioinformatic studies and meta-analyzes that could significantly impact oncological research [36].

3.4. Process of genome data analysis

3.4.1. In a nutshell, what is big data?

Gene expression data are characterized by three facts: (1) they are of high dimensionality (over a thousand genes), (2) the public data are relatively small, with just a few hundred samples available, and (3) the expression of genes in cancerous and non-cancer tissues needs to be distinguished, resulting in numerous genes irrelevant to cancer classification and analysis [37]. To leverage such types of data and develop an accurate classifier based on them, researchers proposed FS and/or dimensionality reduction. Various methods have been used for classification, but recently, DL has been explored because it can handle raw and high-dimensional data [37].

3.4.2. High dimensionality

Pirooznia *et al*. have adapted and tested many classical machine learning methods for transcriptomic applications, including linear models, quadratic models, support vector machines, RFs, and boosting [38]. The accuracy and robustness of medical models can still be a problem [39], even though these methods produce promising results. It is challenging to model gene expression data due to its high dimensionality, the lack of examples for training, overfitting during training, and the lack of robustness [14]. With the help of a stacked sparse autoencoder (AE), meaningful features can be extracted efficiently from high-dimensional data using machine learning approaches like AE [40].

The 11-tumor database is a widely used collection of gene microarrays for studying cancer [41]. The destiny of high-dimensionality is evident in this dataset due to the high number of features and the low capacity of the registers. It suggests that many strategies may be used in the majority of investigations, such as FS algorithms [42-45], dimension reduction [46], clustering methods [47-49], and pre-processing techniques [50], among others.

With most classifiers, gene selection is a preliminary step since gene expression data have a high dimensionality, making classification difficult. By filtering irrelevant features, it improves time complexity and classification accuracy. Although existing "feature selection algorithms" are scalable and generalized, they may not be able to perform accurately on new datasets because of constraints on scalability and generalization. Deep neural networks (DNNs) can help automate the extraction of features and construct generalized and scalable classifiers in such a scenario. As part of this paper, a DL-based genetic algorithm was used with DL to improve cancer classification [51].

3.4.3. FS and dimension reduction

Informative FS is essential to the retrieval of biological data from gene expression data in higher dimensions. An artificial intelligence (AI) model based on DL was developed in a recent study to detect prostate cancer using gene expression data from Microarrays. This technique selected the optimal set of features using an FS technique based on chaotic invasive weed optimization. Our next research article identified causal genes of ovarian cancer using gene features from genomics and pathway annotations. A graph convolutional network was used to construct the gene feature based on gene features and network topology [52]. Graph neural networks (GNNs) also eliminated sparse features from binary fingerprints by extracting graph-level representations containing 2D structures and saving them into feature vectors. A study analyzing lung cancer gene expression profiles in whole blood noted that the dataset were highly imbalanced. AAM and European males had disparities in lung cancer development, which has been addressed using (a) LIMMA, which employs statistical methods, and (b) concrete AE, which utilizes unsupervised algorithms and DL. Several authors have developed an architecture that analyzed whole-transcriptome gene expressions to identify complex genetic alterations that drive cancer progression and complex gene expressions. The architecture is called "Gene Expression Network" (GeneXNet). They used the design of our network to show how it is possible to create an overall end-to-end learning technique for detecting different cancers without selecting genome-wide features [2,9]. A non-linear model may be fitted to choose features. In the case of 50 genes (PAM50), domain experts developed a more autonomous alternative. A computation statistical analysis of the gene set showed that the first 50 features of the PAM50 genes, represented about 1 – 2% of all data. It indicated the importance of the PAM50 genes for fitting the model.

3.5. DL methodologies

Since DL models are better at capturing complicated patterns and correlations within highly-dimensional data, they should be used instead of fundamental models of gene expression analysis for cancer diagnosis, prognosis, and classification. DL models, such as deep convolutional neural networks (CNNs), deep long-short-term memory (LSTM) models, and artificial neural networks with multiple layers, have shown superior performance in handling intricate biological data and extracting meaningful features for accurate predictions [53-55]. This ability to automatically build hierarchical data representations is a significant strength of DL models. It allows them to find hidden relationships and patterns that would have not been obvious using more conventional machine learning techniques. By leveraging multiple layers of hidden units, DL models can effectively capture intricate relationships within gene expression data, enhancing predictive accuracy and robustness in cancer diagnosis and prognosis [55].

Moreover, DL models excel in handling large-scale and highly dimensional datasets, such as gene expression profiles, by automatically extracting relevant features and patterns without requiring manual feature engineering. This capability is conducive to cancer research, where the complexity and heterogeneity of biological data necessitate sophisticated modeling techniques to uncover subtle molecular signatures associated with different cancer types and subtypes [56,57]. DL models offer flexibility and scalability in accommodating diverse data types and modalities, including gene expression data, imaging data, and clinical parameters. By leveraging DL algorithms, researchers can integrate multi-omics data sources and develop comprehensive models that capture the

multidimensional nature of cancer biology, leading to the development of more accurate and personalized diagnostic and prognostic tools [58]. In summary, the rationale for utilizing DL algorithms in gene expression analysis for cancer diagnosis and prognosis stems from their ability to handle complex, high-dimension data, automatically learn hierarchical representations, and extract meaningful features without manual intervention. By harnessing the power of DL models, researchers can gain valuable insights from biological data, improve predictive accuracy, and advance precision medicine in oncology.

3.5.1. LSTM

LSTM uses a much larger memory structure to approximate a non-linear function for prediction. As the prior function for features is represented by this prediction model, it plays a crucial role in learning how the data are distributed. A very complex classification task can be learned using LSTM, which has also helped predict breast cancer subcategories. Nevertheless, the generalization of the learning is much more complicated than the usual machine learning techniques since more data and regularization are required. The initialization condition problem can be solved by extracting practical components from the feature space [59]. LSTMs were also developed by Hochreiter and Schimdhuber [60], and are used primarily for voice recognition. Because of using a cell as a memory unit, LSTMs could keep their value for an extended period and thus remember the value they computed recently. Cells, also known as memory units, contain three gates that control data movement within [6].

3.5.2. Deep recurrent neural networks (DRNN)

With the DRNN classifier, survival can be predicted in a shorter period, so it is an effective tool for survival prediction [61]. The DRNN classifier records its dynamic time series and hidden layer through the nodes' directional connections within the hidden layer. The configuration of DRNN differs from that of a feed-forward network using feed-forward and feedback associations between the internal processing elements, which record input sequences at different times. DRNN achieves a more robust transformation with technical indicators by combining various non-linear layers associated with time, allowing it to improve prediction capacity. This method predicts outputs based on information from random sequences and historical data at lower phases. In addition, the data sequence is efficiently processed, followed by applying the output. Using feature learning, inputs are mapped to hidden states, and then hidden states are mapped to output series [7].

An approach using CNNs and bidirectional gated recurrent units (BiGRUs) was proposed to reduce the dimensionality of gene expression data and remove irrelevant components. DCGN initially used synthetic minority over-sampling to ensure data parity. In contrast to CNNs, which struggle with high-dimension data and extract critical local properties, BiGRUs could examine deep features while preserving their information. The DCGN has small sample sizes and sparse features. Hence, a combination of neural networks is needed to capture them [1]. A quantum-inspired immune clone optimization algorithm (QICO) was used to optimize the hidden neurons of the optimized RNN using a FS and classification algorithm (QICO). Using the constraints of 5 and 25 hidden neurons, QICO schemes tune the hidden neurons to boost performance [62].

3.5.3. Transfer learning

Transfer learning is often recommended for dealing with gene expression data with small training sets and high dimensions [63]. Transfer takes information from one model (source) and transfers it to another. Any classification task involving common visual or textual patterns is commonly solved using transfer learning [63] in image analysis and natural language processing. It is possible to use a large source dataset and a target dataset to develop these classes. Researchers evaluated the performance of a DL approach based on an exhaustive set of experiments for predicting cancer [14]. It has been shown that gene expression data could be transferred between cell lines and patient data, as well as pan-cancer and specific cancers. This approach may help develop accurate models in rare cancer cases where large datasets are unavailable.

In another study, a system of end-to-end learning was developed using multiple tissue samples representing a variety of cancerous tumors spreading across various organs. Total RNA sequencing was used in the deep CNN learning model to quantify gene expression across the whole transcriptome [2,10].

3.5.4. DNN

By utilizing the multi-layer feature of DL [64], multi-layer DNN can effectively explore hierarchical data representation [37]. Due to these properties, DNNs showed outstanding performance in cancer classification. As one of the simplest types of multi-layer neural networks, DNNs can handle AE, stacked AE, deep belief network, and Boltzmann machines with many advantages, such as utilizing perceptrons.

3.5.5. Adversarial networks

The researchers provided a DL model for cancer detection by employing gene expression patterns from the whole transcriptome [2,9,10]. They designed a novel CNN topology to identify cancer-causing genetic mutations in

whole transcriptomes. To this end, genomic signatures were learned across many tissue types without first having to choose gene characteristics. To cope with complex gene expression and limited training samples, a new CNN architecture called GeneXNet uses multi-layer blocks called GeneXNet blocks. This network uses fewer input samples and is highly accurate. Two residual learning networks are developed by combining deep CNNs and densely connected convolutional networks. On the test dataset, the SROC accuracy was 98.93%, and the area under the curve (AUC) was 0.99 based on the block architecture combined with two different learning sub-blocks.

A deep generative machine learning algorithm, DeepCancer, identifies non-labeled features in microarray data [65]. DeepCancer has a DL component since it uses Generative Adversarial Networks to classify tissue samples. The model merges traditional classifiers with a new model to identify cancerous and non-cancerous tissues. A standard uniform distribution is used to collect an example of a noise mini-batch, and a standard distribution is used to collect an example of a data mini-batch. The gradients can be calculated using an optimization function. The training of discriminators should be frozen. They applied a mini-batch of noise samples from the standard uniform distribution and then performed stochastic gradient descent on G. Using two clinical datasets, they tested the model. The F-score was 70%, the precision was 55%, and the recall was 100%.

Five models were tested on five sets of data from the Omnibus library [37]. The purpose of this study was to define deep feed-forward neural networks. They used a multilayer perceptron (MLP) feed-forward neural network to receive the gene expression values of each sample. They also applied a dropout penalty to three dropout layers and four fully-connected layers to prevent over-fitting. The network also had seven hidden layers. A Softmax classifier also assigned the output features of the seventh hidden layer, and the input layer then applied a regularization parameter to the data. Finally, layers were used for non-linear relative and tangent hyperbolic (tanh) functions. About 100%, 85%, 100%, 96%, and 100% accuracy were achieved in five datasets, respectively [37].

3.5.6. AE

Tapak *et al*. used a hierarchical clustering technique to identify high-risk and low-risk groups that relied on just 100 AE features retrieved using a deep-learning neural network [66]. Gene profiles linked to different forms of oral cancer were found using the Cox regression model and the supervised RF technique. Statistical evaluations of the GEO dataset yielded positive results.

A study [59] used gene-subcategory interactions (GSIAR) to regularize breast cancer data. To reach the best balance in representation, GSIAR combined regeneration and sub-categorical capabilities. Using GSIAR, genes associated with human disorders were heuristically selected according to their association with them. A deep computational architecture was used to construct prediction models based on cleansing, modeling, and analysis of statistical data. As a result, they developed a concept of selection and analysis and a prediction model, resulting in an F Score of 83.3% at the American Joint Committee on Cancer.

3.5.7. CNN

DL is achieved using CNNs with BiGRUs [1]. By learning features from gene expression data, DCGN reduces non-linear dimensionality and removes irrelevant factors. It uses gene expression data to identify cancer subtypes. Using the Synthetic Minority Oversampling Technique algorithm, samples are balanced, and enhanced data are obtained using the first DCGN module. Three modules make up the program. A second part of the process involves normalizing the features. The third part of the process comprises capturing critical gene expression data features during training. This research used breast cancer gene expression datasets to predict multi-classifications using the outputs of the feature learning module. Classification losses were calculated by comparing false labels with actual labels. The BLCA-CIT-Curie database was used to analyze these results, and they showed accuracy, precision, recall, and specificity values of 98.5%, 98.7%, 98.5%, and 98.5%, respectively [1].

The DL techniques of MLP, two-dimensional, and one-dimensional CNNs were compared by Majumder *et al*. Features for colon, pancreatic, breast, and lung malignancies were selected using Analysis of variance (ANOVA) and information gain. ANOVA and 1DCNN were used to select features. 1DCNN and IG FS yielded 100% sensitivity, 100% precision, and 100% accuracy for the classification of lung cancer [67].

Almarzouki used bone marrow PC gene expression data [8] to present the artificial Bee Colony features selection method. Tumors were unlabeled and then classified using CNNs. Cross-validation of k-folds was also deployed for CNNs using the Kaggle genomic dataset, including lung, kidney, and brain cancer datasets. The precision, sensitivity, accuracy, and F-score rates were determined as 86%, 86%, 98.97%, and 86%, respectively [8].

Using a dual convergence architecture, [3] information about mutations, gene expression, and drug structures was separated. An algorithm was created by combining the data from both models to build a single prediction model. GNNs and CNNs were used to extract independent features from genomic signatures and molecular graphs, which were then merged. The genomics of cancer drug sensitivity (GDSC) and cancer cell lines encyclopedia datasets were used to study the

GDSC. Neither F-score nor sensitivity, specificity, or accuracy were provided.

Zhao *et al*. used a method to predict clinical outcomes based on tumor genomic profiles without gene filtering, according to [16]. This method used CNNs with stationary wavelet transforms (SWTs). SVM, logistic regression, and RF were all included in the proposed SWT-CNN to increase prediction accuracy. Before training the CNN, they used Cox proportional-hazards regression to choose predictive markers by ranking the genes with the ranking values. Databases, such as TCGA, were utilized for the research. Three-year survival was predicted with an AUC of 0.664, whereas the tumor stage was predicted, with an AUC being 0.661 on average.

Shah *et al.* [70] proposed a hybrid LS-CNN DL model for classifying cancer data. The data were cleaned of categorical row and column values, missing values, and numeric features. LS then assigned a Laplacian score to each feature. LS used nearest-neighbor graphs to model local geometric structures. LS then selected features related to this graph structure. A CNN was then applied to classify the data. Datasets indicated an average accuracy of 97.93%.

According to Tabares-Soto *et al*. [68], the accessible "Tumors database" (https://github.com/simonorozcoarias/ML_DL_microArrays/blob/master/data11tumors2.csv) was used. The best accuracy was 94.43% using k-fold cross-validation. A CNN tuning did not significantly improve the accuracy.

Table 1 summarizes the literature about the performance of DL approaches used in different cancer types, along with their corresponding gene expression database.

3.6. Ethical concerns and regulations regarding AI applications in oncology

Implementing AI in health-care settings, particularly oncology offers a transformative pathway to enhancing patient care and outcomes. However, this integration raises significant ethical implications and necessitates regulatory considerations to ensure responsible and effective deployment [73]. The following are critical ethical considerations: transparency, informed consent, impartiality, and compliance with regulations such as the European Union's AI Act and HIPAA [74]. For AI to reach its full potential in cancer management and healthcare in general, it is essential to address the ethic, technological, and regulatory issues [73]. The ethical considerations in implementing AI in health-care extend to data privacy concerns, regulatory hurdles, and ensuring fairness and equity in deploying AI and robotics [75]. When incorporating AI and machine learning into clinical practice, stakeholders must navigate ethical dilemmas, data privacy, and regulatory landscapes [76]. Together, stakeholders can raise awareness about potential ethical problems and emphasize responsible AI development to create a future in which AI enhances health-care delivery in a way that is both ethical and beneficial to patients [77]. Regulatory frameworks are crucial to safeguarding patient rights and promoting the ethical use of AI in healthcare [78]. Establishing normative standards and evaluation guidelines for AI implementation in health-care requires collaboration among regulatory agencies and health-care institutions [79].

Furthermore, interdisciplinary collaborations are advocated for establishing ethical guidelines and ensuring responsible AI use in healthcare [80]. The successful integration of AI in healthcare, especially in oncology, depends on thoroughly considering ethical implications and regulatory frameworks. Transparency, fairness, data privacy, and adherence to regulations are paramount to exploiting the full potential of AI while safeguarding patients' well-being and trust in the health-care system.

3.7. Scalability of the DL approaches

Scalability is a crucial aspect when considering the implementation of DL approaches in clinical settings, particularly in terms of computational costs and real-time applicability. DL models, such as CNN, AEs, Adversarial Networks, DNN, Transfer Learning, DRNN, and LSTM networks, have shown promise in revolutionizing health-care applications [56,81,82]. These models offer a scalable and data-driven approach to understanding complex systems and extracting valuable insights from large datasets [56,81,82]. In clinical settings, the scalability of DL approaches is essential to efficiently handling vast amounts of health-care data. DL models have demonstrated the ability to learn patterns within data and make accurate predictions, such as in disease diagnosis and image processing tasks [83-85]. However, the scalability of these models comes at computational costs that must be carefully ensured for real-time applicability [86,87]. Scientists have suggested new ideas to tackle the computational difficulties linked to DL in the health-care field. One suggestion is to create scalable deep-learning systems that can learn important clinical traits by utilizing historical data from medical ontologies [88,89].

Advancements in DL accelerators and edge computing have aimed to improve the efficiency and real-time applicability of DL models in health-care applications. Furthermore, to make health-care data management more secure and scalable, people have looked at ways to combine DL with other technologies, such as blockchain. These hybrid approaches offer promising solutions to address scalability issues while ensuring data security and integrity in clinical settings. While DL approaches hold great potential for transforming healthcare, including oncology, their scalability in terms of

Table 1. A list of deep learning approaches used for predicting cancer and its subtypes using the gene datasets

Study	Method	Dataset	Accuracy	Area under the curve	F-score	Precision	Recall	Specificity
[9]	Two residual learning networks, i.e., deep and complex CNNs and densely connected convolutional networks	11,093 human samples for mRNA gene expression, 60,483 genes	98.93	0.99	-	-	-	-
[65]	DeepCancer	GSE45584 breast cancer dataset and prostate cancer dataset [69]	-	-	70	55	100	-
[37]	Deep learning recursive feature elimination	Leukemia cancer, GSE15061, 54613 genes	90	-	-	-	-	-
	Deep learning univariate feature selection	Inflammatory breast cancer, GSE45581, 40991 genes	85	-	-	-	-	-
	Deep learning recursive feature elimination	Lung cancer, GSE2088, 40368 genes	100	-	-	-	-	-
	Deep learning univariate feature selection	Lung cancer, GSE2088, 40368 genes	100	-	-	-	-	-
	Deep learning univariate feature selection	Bladder cancer, GSE31189, 54675 genes	96	-	-	-	-	-
	Deep learning recursive feature elimination	Thyroid cancer, GSE82208, 54671 genes	100	-	-	-	-	-
[66]	AutoEncoder, Cox proportional hazards model, and a supervised random forest	GSE26549	91.6	-	-	-	81.4	96.6
[59]	Dual-phase deep learning (deep representation learning)	American joint committee on cancer stage	-	-	83.3	-	-	-
[67]	Multilayer perceptron IG (information gain)	Colon	84	-	87	87	87	-
	1DCNN IG	Pancreas	91	-	93	87	100	-
	1DCNN IG	Breast	94	-	92	86	100	-
	1DCNN IG	Lung	100	-	100	100	100	-
[8]	Artificial bee colony, CNNs	Kaggle gene expression dataset	98.97		86	86	86	-
[3]	Graph Neural Networks and CNN	SWnet to genomics of cancer drug sensitivity and cancer cell lines encyclopedia	-	-	-	-	-	-
[16]	SWT-CNN	TCGA tumor stage prediction	-	0.661	-	-	-	-
	SWT-CNN	TCGA 3-year overall survival prediction	-	0.664	-	-	-	-
[70]	CNN based on Laplacian score	AML, CNS, MLL, SRBCT, Colon, Arcene, DLBCL, Leukemia, Prostate, Brain	97.93	-	-	-	-	-
[68]	Using k-fold cross-validation, a CNN is applied	11 Tumors database	94.43	-	-	-	-	-
[71]	PathDeep	Gene expression omnibus microarray	99.1	-	-	-	-	-
[4]	Wasserstein generative adversarial network	TCGA	-	0.98	-	-	-	-
[6]	Stacked ensemble model	Five benchmark datasets	98	-	-	-	-	-
[14]	NN	TCGA	98.05	-	-	-	-	-
[52]	XGPG	GWAS and eQTL	-	0.7541	-	-	-	-
[39]	AIFSDL-PCD, chaotic invasive weed optimization, deep neural network	Open access dataset, including 52 prostate tumors and 50 normal tissues with 2135 genes.	97.47	-	97.58	96.87	97.75	97.26
[72]	proposed CNN DeepGX	TCGA	95.65	-	94.45	95.55	95.69	-
[51]	DLFCC	Leukemia, Lung, Prostate, Colorectal	86.25	-	-	-	87.75	67.5
[7]	Deep recurrent neural networks -based CHHPO	PANcancer	52.5	-	-	-	-	-

(Cont'd...)

Table 1. ***(Continued)***

Study	Method	Dataset	Accuracy	Area under the curve	F-score	Precision	Recall	Specificity
[1]	DCGN	BRCA	94.8	-	-	96.8	96.7	97
	DCGN	BLCA	94.2	-	-	94.5	94.2	94.2
	DCGN	BLCA, Lund	93	-	-	93.7	93.4	93.4
	DCGN	BLCA TCGA	98.2	-	-	98.4	98.2	98.2
	DCGN	BLCA-CIT-Curie	98.5	-	-	98.7	98.5	98.5
[62]	QICO-RNN	Colon cancer, defused B-cell Lymphoma, Leukemia, Wisconsin Diagnostic Breast Cancer and Wisconsin Breast Cancer Data"	Not available numerically (over 95)	-	Not available numerically (over 80)	Not available numerically (over 80)	Not available numerically (over 80)	Not available numerically (over 80)

CNN: Convolutional neural network, SWT: Stationary wavelet transforms, -: No values for the corresponding metrics, TCGA: The cancer genome atlas

computational costs and real-time applicability is a critical consideration. By developing scalable frameworks, leveraging innovative technologies, and optimizing computational resources, DL can be effectively implemented in clinical settings to improve patient care and outcomes.

DL approaches, such as those mentioned above, provide powerful tools for analyzing complex health-care data in clinical settings. However, the computational costs and real-time applicability of these models are critical considerations for their successful implementation [90,91]. The computational costs associated with DL models can be significant due to the complexity of the algorithms and the large amounts of data involved. While DL models excel at learning intricate patterns within data, this efficacy often comes at the expense of increased computational complexity [91]. Managing these computational costs is essential to ensuring the practicality and efficiency of deploying DL approaches in real-time clinical settings [90]. To address the computational challenges, researchers have explored innovative solutions to enhance the scalability and real-time applicability of DL models in healthcare. For instance, integrating DL with edge computing technologies has shown promise in improving the efficiency and speed of data processing, enabling real-time analysis and decision-making [90].

In addition, advancements in DL accelerators and hardware have aimed to optimize computational resources and enhance the real-time performance of DL models in clinical applications [90]. Moreover, the development of hybrid DL models that leverage blockchain technology has been proposed to enhance the scalability, security, and real-time processing of health-care data [88]. By combining DL with blockchain, researchers aim to address computational costs while ensuring data integrity and security in real-time health-care applications. So, while DL approaches offer immense potential for revolutionizing healthcare, including oncology, addressing these models' computational costs and real-time applicability is crucial to their successful integration into clinical settings. DL can improve patient-care outcomes in real-time clinical decision-making by leveraging innovative technologies, optimizing computational resources, and developing scalable frameworks.

3.8. Case studies

In oncological practice, DL models have shown significant effectiveness in assisting with cancer diagnosis and prognosis, leading to notable advancements in patient care. Various studies have highlighted the transformative impact of these models on improving outcomes for cancer patients. One example was a study focusing on DL -based cancer survival prognosis using RNA-seq data. The researchers developed and assessed three distinct deep-learning models for cancer prognosis, demonstrating that these models had the potential to offer precise and personalized survival predictions based on genomic data [92]. This study illustrated how DL methods could use molecular data to enhance the accuracy of cancer prognosis. In another study, Lai *et al.* used DL to predict the OS of non-small cell lung cancer (NSCLC) patients by integrating microarray and clinical data. The researchers illustrated the effectiveness of DL by combining diverse data sources to enhance prognostic predictions for different NSCLC subtypes, such as adenocarcinoma and squamous cell carcinoma [93]. This research underscores the value of DL in integrating multi-omics data to improve cancer prognosis.

The research on the use of AI methods for breast cancer diagnosis and prognosis emphasized that improved accuracy, efficiency, and reliability could be achieved through DL models. By employing CNN, a study highlighted the potential for DL in improving breast cancer diagnosis and prognosis, underscoring the significance of these models in clinical decision-making [94]. The study additionally explored the application of DL in cancer prognosis prediction, highlighting the potential of DL models in utilizing multi-omics data for enhanced prognostic accuracy. By incorporating genomics, transcriptomics, and clinical information, DL models offer a promising approach to improving cancer prognosis and

survival prediction (Zhu *et al*., 2020). These case studies showed the substantial impact of DL models in oncology, showcasing their ability to revolutionize cancer diagnosis and prognosis by leveraging diverse data sources and enhancing predictive accuracy. Integrating DL approaches into clinical practice allows health-care professionals to make more informed decisions, ultimately improving patient outcomes and personalized treatment strategies.

3.9. Potent interpretability of the DL models in cancer diagnosis

It is essential to explore techniques that improve model transparency and trustworthiness to enhance the interpretability of DL models in critical fields like cancer diagnosis. Several studies have delved into this area, addressing challenges and looking for opportunities to make DL models more interpretable.

Meng *et al*. studied the interpretability and fairness evaluation of DL models for mortality prediction using the Medical Information Mart for Intensive Care-IV dataset. They highlighted the importance of interpretability methods in identifying critical features for mortality prediction and addressing fairness concerns in model prediction [95]. This study underscored the significance of transparent and fair DL models in health-care applications. In another study, Miotto *et al.* discussed the challenges in DL applications in healthcare. They emphasized the need to develop interpretable architectures to enhance the understanding of DL models. By bridging the gap between complex DL models and human interpretability, researchers could improve the trustworthiness of these models in critical applications like cancer diagnosis [96].

Furthermore, classification of lung and colon cancer has also been explored by using medical imaging and the study demonstrated the advantage of machine learning models in providing better interpretability through feature engineering. The study compared the interpretability of machine learning models with the black-box nature of DL models, emphasizing the importance of improving the transparency of DL approaches [97].

Researchers have proposed innovative solutions to address the lack of interpretability in DL models. For example, attention mechanisms were investigated to enhance the interpretability in DL models, providing insights into the features influencing predictions [98]. The challenges posed by the poor interpretability of conventional CNN models in cancer prognosis prediction were also addressed, emphasizing the need for more transparent and interpretable models [99]. In conclusion, improving the interpretability of DL models for cancer diagnosis is crucial for fostering trust and understanding in clinical decision-making. By developing transparent and interpretable DL architectures, researchers can enhance the reliability and acceptance of these models in critical health-care applications.

4. SEMINAL WORKS

To ensure a comprehensive discussion of DL methods for cancer diagnosis, it is crucial to integrate foundational theories with innovative research. By including seminal works alongside recent studies, the manuscript offers a thorough overview of the evolution and current status of DL applications in cancer diagnosis. One seminal work is the study conducted by Wang *et al*. [100], who used DL to analyze images from lung cancer pathologies and showed that the future of DL in cancer diagnosis presents both obstacles and opportunities. This work added to our fundamental knowledge by shedding light on how DL algorithms were used for lung cancer detection and prognosis. Another essential study to consider is the work by Verma [101], which examined the application of interpretable DL models in health-care for disease diagnosis. This research emphasized the significance of model interpretability in developing reliable and trustworthy DL systems for medical applications.

Furthermore, the study by Wei [102] on radiomics, DL, and early diagnosis in oncology can serve as a foundational reference. This work discussed the role of medical imaging in cancer detection and highlighted the importance of DL approaches in early diagnosis and prediction of treatment response. By incorporating seminal works alongside recent studies, this review presents a well-rounded discussion on the evolution, challenges, and advancements in DL methods for cancer diagnosis. This approach helps readers understand the field comprehensively, ranging from foundational principles to cutting-edge applications.

5. LIMITATIONS AND FUTURE DIRECTIONS

In the realm of DL approaches for gene expression analysis for cancer diagnosis, prognosis, and classification, it is crucial to assess the limitations of current studies. By scrutinizing the strengths and weaknesses of existing studies, we can pinpoint areas for improvement and innovation in utilizing DL models in oncology.

A significant limitation of current studies is the lack of interpretability and transparency in DL models, particularly in gene expression analysis for cancer diagnosis. The fundamental processes underlying the predictions of many DL models are sometimes seen as “black boxes” because of this limitation. As a result, DL methods may not be suitable for therapeutic settings where trustworthiness and openness are more important [101]. Although DL models have

demonstrated potential in survival prediction and analysis of histopathological images of various cancer types, challenges may arise concerning model generalizability and robustness across diverse patient populations and datasets. Variations in data quality, sample sizes, and feature representation can influence the performance and dependability of DL models in real-world clinical settings [103]. The scalability and computational expenses associated with DL models for gene expression analysis for cancer diagnosis can present obstacles, particularly when applying these models in resource-limited health-care environments. Enhancing model efficiency, reducing computational burdens, and ensuring real-time applicability are critical for putting DL approaches into oncological applications [102].

This review endeavored to tackle these limitations by offering a comprehensive analysis of the current landscape of DL applications in gene expression analysis for cancer diagnosis. By integrating seminal works with recent studies, the review provided a well-rounded perspective on the progression of DL methods in oncology. By critically assessing the gaps and challenges to existing studies, this review laid the groundwork for future advancements in creating interpretable, robust, and scalable deep-learning models for cancer diagnosis and prognostic evaluation of cancer. Future directions in applying DL for gene expression analysis for cancer diagnosis, prognostication, and classification are crucial for advancing precision medicine and improving patient outcomes. By synthesizing insights from current research and identifying emerging trends, researchers can shape the trajectory of DL applications in oncology. One potential future direction is integrating multi-omics data to enhance the predictive power of DL models for cancer diagnosis and prognostic prediction. By combining genomics, transcriptomics, proteomics, and other molecular data types, researchers can develop comprehensive models that capture the complexity of cancer biology and improve patient stratification and treatment selection [104].

Exploring explainable AI techniques in DL models for cancer diagnosis and prognostication is a promising avenue for future research. Enhancing the interpretability and transparency of DL models can foster trust among clinicians and patients, facilitating the adoption of AI-driven decision support systems in clinical practice [105]. Another future direction involves the development of personalized deep-learning models for cancer treatment selection. By leveraging patient-specific data, such as genetic profiles, imaging data, and clinical parameters, researchers can tailor treatment recommendations using machine learning algorithms, leading to more effective and individualized cancer therapies [106].

The advancement of DL models for early cancer detection and monitoring holds significant promise. DL algorithms may be trained to identify cancer earlier with more precision and sensitivity using new imaging modalities, such as digital pathology and radiomics. Early detection and timely treatment will help attain better patient outcomes.

By investigating federated learning methodologies, we can ensure the privacy and security of patient data while facilitating cross-institutional model training. Robust and generalizable DL models may be trained using federated learning, aggregating information from multiple datasets without centralized data sharing. It opens the door to scalable and privacy-preserving AI applications in cancer research. In conclusion, future research directions in DL for gene expression analysis for cancer diagnosis encompass integrating multi-omics data, explainable AI techniques, personalized treatment selection, early detection strategies, and federated learning approaches. By embracing these innovative pathways, researchers can take the field of oncology toward more precise, efficient, and patient-centered cancer care.

6. CONCLUSIONS

This review explicitly shows that DL approaches have immense potential for revolutionizing gene expression analysis for cancer diagnosis, prognostication, and classification. The various DL methodologies, including LSTM Networks, DRNN, CNN, and AEs, have experienced significant advancements in handling the complex, high-dimensional nature of genomic data.

These approaches have been proven effective in extracting meaningful features from gene expression profiles, reducing dimensionality, and improving the accuracy of cancer classification and prognostic assessment. DL models' ability to automatically learn hierarchical representations from raw data has enabled researchers to uncover subtle patterns and relationships within gene expression datasets that may have been subtle using traditional analytical methods. However, the review also highlights several challenges and limitations faced by current DL applications in oncology. These include interpretability, scalability, and the need for large datasets to train robust models. The ethical concerns and regulatory considerations surrounding the use of AI in health-care settings, particularly in sensitive areas, such as cancer diagnosis, underscore the importance of responsible development and implementation of these technologies. This review also points to the directions for future research and development. Integrating multi-omics data, exploring explainable AI techniques, and developing personalized deep-learning models for treatment selection can significantly enhance these approaches' clinical utility. The potential for early cancer detection and monitoring using advanced imaging modalities and DL algorithms also presents exciting possibilities for improving patient outcomes.

In conclusion, while DL approaches have made significant strides in enhancing our understanding and analysis of gene expression in cancer, substantial room remains for growth and refinement. As researchers continue to address current limitations and explore new frontiers, integrating DL into oncology can dramatically improve cancer diagnosis, prognostication, and treatment efficacy, ultimately leading to better patient care and outcomes. The ongoing collaboration between computational scientists, biologists, and clinicians will be crucial in realizing the full potential of these powerful analytical tools in the fight against cancer.

ACKNOWLEDGMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Babak Sokouti

Data curation: All authors

Writing–original draft: All authors

Writing–review and editing: All authors

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA

Not applicable.

REFERENCES

1. Shen J, Shi J, Luo J, *et al*. Deep learning approach for cancer subtype classification using high-dimensional gene expression data. *BMC Bioinformatics*. 2022;23(1):430. doi: 10.1186/s12859-022-04980-9
2. Khorshed T, Moustafa MN, Rafea A. Deep learning for multi-tissue cancer classification of gene expressions (GeneXNet). *IEEE Access*. 2020;8:90615-90629. doi: 10.1109/ACCESS.2020.2992907
3. Zuo Z, Wang P, Chen X, Tian L, Ge H, Qian D. SWnet: A deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC Bioinformatics*. 2021;22(1):434. doi: 10.1186/s12859-021-04352-9
4. Xiao Y, Wu J, Lin Z. Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data. *Comput Biol Med*. 2021;135:104540. doi: 10.1016/j.compbiomed.2021.104540
5. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: A new generation of clinical biomarkers. *Br J Cancer*. 2021;124(4):686-696. doi: 10.1038/s41416-020-01122-x
6. Gupta S, Gupta MK, Shabaz M, Sharma A. Deep learning techniques for cancer classification using microarray gene expression data. *Front Physiol*. 2022;13:952709. doi: 10.3389/fphys.2022.952709
7. Majji R, Maram B, Rajeswari R. Chronological horse herd optimization-based gene selection with deep learning towards survival prediction using PAN-Cancer gene-expression data. *Biomed Signal Process Control*. 2023;84:104696. doi: 10.1016/j.bspc.2023.104696
8. Almarzouki HZ. Deep-learning-based cancer profiles classification using gene expression data profile. *J Healthc Eng*. 2022;2022:4715998. doi: 10.1155/2022/4715998
9. Khorshed T, Moustafa MN, Rafea A, editors. Multi-tissue Cancer Classification of Gene Expressions using Deep Learning. In: *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications* (*BigDataService*). IEEE; 2020. doi: 10.1109/BigDataService49289.2020.00027
10. Khorshed T, Moustafa MN, Rafea A, editors. Learning and Visualizing Genomic Signatures of Cancer Tumors Using Deep Neural Networks. In: *2020 International Joint Conference on Neural Network*s (*IJCNN)*. IEEE; 2020. doi: 10.1109/IJCNN48605.2020.9207368
11. Mallik S, Seth S, Bhadra T, Zhao Z. A Linear regression and deep learning approach for detecting reliable genetic alterations in cancer using DNA Methylation and Gene expression data. *Genes* (*Basel*). 2020;11(8):931. doi: 10.3390/genes11080931
12. Rezaee K, Jeon G, Khosravi MR, Attar HH, Sabzevari A. Deep learning-based microarray cancer classification and ensemble gene selection approach. *IET Syst Biol*. 2022;16(3-4):120-131. doi: 10.1049/syb2.12044
13. Wang S, Zhang H, Liu Z, Liu Y. A novel deep learning method to predict lung cancer long-term survival with biological knowledge incorporated gene expression images and clinical data. *Front Genet*. 2022;13:800853. doi: 10.3389/fgene.2022.800853
14. Hanczar B, Bourgeais V, Zehraoui F. Assessment of deep learning and transfer learning for cancer prediction based on gene expression data. *BMC Bioinformatics*. 2022;23(1):262. doi: 10.1186/s12859-022-04807-7
15. Liu S, Yao W. Prediction of lung cancer using gene expression and deep learning with KL divergence gene selection. *BMC Bioinformatics*. 2022;23(1):175.

doi: 10.1186/s12859-022-04689-9
16. Zhao Y, Zhou Y, Liu Y, *et al*. Uncovering the prognostic gene signatures for the improvement of risk stratification in cancers by using deep learning algorithm coupled with wavelet transform. *BMC Bioinformatics*. 2020;21(1):195. doi: 10.1186/s12859-020-03544-z
17. Orrico KB. Basic concepts of cancer genetics and receptor tyrosine kinase inhibition for pharmacists. A narrative review. *J Oncol Pharm Pract*. 2023;29(5):1187-1195. doi: 10.1177/10781552231153814
18. Pettini F, Visibelli A, Cicaloni V, Iovinelli D, Spiga O. Multi-omics model applied to cancer genetics. *Int J Mol Sci*. 2021;22(11):5751. doi: 10.3390/ijms22115751
19. Ferguson LR, Chen H, Collins AR, *et al*. Genomic instability in human cancer: Molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. *Semin Cancer Biol*. 2015;35(Suppl):S5-S24. doi: 10.1016/j.semcancer.2015.03.005
20. Hesson LB, Hitchins MP, Ward RL. Epimutations and cancer predisposition: Importance and mechanisms. *Curr Opin Genet Dev*. 2010;20(3):290-298. doi: 10.1016/j.gde.2010.02.005
21. Hassanpour SH, Dehghani M. Review of cancer from perspective of molecular. *J Cancer Res Pract*. 2017;4(4):127-129. doi: 10.1016/j.jcrpr.2017.07.001
22. Seligson DB, Horvath S, Shi T, *et al*. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature*. 2005;435(7046):1262-1266. doi: 10.1038/nature03672
23. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19(11):1423-1437. doi: 10.1038/nm.3394
24. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. *Nat Genet*. 2018;50(9):1212-1218. doi: 10.1038/s41588-018-0202-0
25. De Anda-Jáuregui G, Hernández-Lemus E. Computational oncology in the multi-omics era: State of the Art. *Front Oncol*. 2020;10:423. doi: 10.3389/fonc.2020.00423
26. Ergin S, Kherad N, Alagoz M. RNA sequencing and its applications in cancer and rare diseases. *Mol Biol Rep*. 2022;49(3):2325-2333. doi: 10.1007/s11033-021-06963-0
27. Mortezaei Z. Computational methods for analyzing RNA-sequencing contaminated samples and its impact on cancer genome studies. *Inform Med Unlocked*. 2022;32:101054. doi: 10.1016/j.imu.2022.101054
28. Sezer A. A review on microarray technology in molecular diagnostics. *Bioeng Stud*. 2021;2(2):7-11. doi: 10.37868/bes.v2i2.id191
29. Cowell JK, Hawthorn L. The application of microarray technology to the analysis of the cancer genome. *Curr Mol Med*. 2007;7(1):103-120. doi: 10.2174/156652407779940387
30. Koch M, Wiese M. Accessing cancer metabolic pathways by the use of microarray technology. *Curr Pharm Des*. 2013;19(4):790-805.
31. Li L, Li L, Sun Q. High expression of cuproptosis-related SLC31A1 gene in relation to unfavorable outcome and deregulated immune cell infiltration in breast cancer: An analysis based on public databases. *BMC Bioinformatics*. 2022;23(1):350. doi: 10.1186/s12859-022-04894-6
32. Zhong Z, Hong M, Chen X, *et al*. Transcriptome analysis reveals the link between lncRNA-mRNA co-expression network and tumor immune microenvironment and overall survival in head and neck squamous cell carcinoma. *BMC Med Genomics*. 2020;13(1):57. doi: 10.1186/s12920-020-0707-0
33. Athar A, Füllgrabe A, George N, *et al*. ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Res*. 2018;47(D1):D711-D715. doi: 10.1093/nar/gky964
34. Rhodes DR, Yu J, Shanker K, *et al*. ONCOMINE: A cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1-6. doi: 10.1016/S1476-5586(04)80047-2
35. Nwosu IO, Piccolo SR. A systematic review of datasets that can help elucidate relationships among gene expression, race, and immunohistochemistry-defined subtypes in breast cancer. *Cancer Biol Ther*. 2021;22(7-9):417-429. doi: 10.1080/15384047.2021.1953902
36. Alameer A, Chicco D. geoCancerPrognosticDatasetsRetriever: A bioinformatics tool to easily identify cancer prognostic datasets on Gene Expression Omnibus (GEO). *Bioinformatics*. 2022;38(6):1761-1763. doi: 10.1093/bioinformatics/btab852
37. Zenbout I, Meshoul S, editors. Advanced Machine Learning Models for Large Scale Gene Expression Analysis in Cancer Classification: Deep Learning Versus Classical Models. In: *Big Data, Cloud and Applications: Third International Conference, BDCA. Revised Selected Papers 3*. Kenitra, Morocco: Springer; 2018.
38. Pirooznia M, Yang JY, Yang MQ, Deng Y. A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*. 2008;9(1):S13. doi: 10.1186/1471-2164-9-S1-S13
39. Alshareef AM, Alsini R, Alsieni M, *et al*. Optimal deep learning enabled prostate cancer detection using microarray gene expression. *J Healthc Eng*. 2022;2022:7364704. doi: 10.1155/2022/7364704
40. Bhonde SB, Prasad JR, editors. Deep Learning Techniques in Cancer Prediction Using Genomic Profiles. In: *2021 6th International Conference for Convergence in Technology* (*I2CT*). Maharashtra, India; 2021. doi: 10.1109/I2CT51068.2021.9417985
41. Su AI, Welsh JB, Sapinoso LM, *et al*. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res*. 2001;61(20):7388-7393.
42. Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A, Benítez JM, Herrera F. A review of microarray datasets and

applied feature selection methods. *Inf Sci*. 2014;282:111-135. doi: 10.1016/j.ins.2014.05.042
43. Wang S, Wei J. Feature selection based on measurement of ability to classify subproblems. *Neurocomputing*. 2017;224:155-165. doi: 10.1016/j.neucom.2016.10.062
44. Han D, Kim J. Unified simultaneous clustering and feature selection for unlabeled and labeled data. *IEEE Trans Neural Netw Learn Syst*. 2018;29(12):6083-6098. doi: 10.1109/TNNLS.2018.2818444
45. Perera K, Chan J, Karunasekera S, editors. Feature Selection for Multiclass Binary data. In: *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, Proceedings, Part III 22*. Springer; 2018. doi: 10.1007/978-3-319-93040-4_5
46. Araújo D, Neto AD, Martins A, Melo J, editors. Comparative Study on Dimension Reduction Techniques for Cluster analysis of Microarray Data. In: *The 2011 International Joint Conference on Neural Networks*. IEEE; 2011. doi: 10.1109/IJCNN.2011.6033447
47. Sardana M, Agrawal RK, editors. A Comparative Study of Clustering Methods for Relevant Gene Selection in Microarray Data. In: *Advances in Computer Science, Engineering and Applications: Proceedings of the Second International Conference on Computer Science, Engineering and Applications* (*ICCSEA 2012*). Vol. 1. New Delhi, India: Springer; 2012.
48. Sirinukunwattana K, Savage RS, Bari MF, Snead DR, Rajpoot NM. Bayesian hierarchical clustering for studying cancer gene expression data with unknown statistics. *PLoS One*. 2013;8(10):e75748. doi: 10.1371/journal.pone.0075748
49. Li J, Liu R, Zhang M, Li Y, editors. Ensemble-based Multi-objective Clustering Algorithms for Gene Expression Data Sets. In: *2017 IEEE Congress on Evolutionary Computation* (*CEC*). IEEE; 2017. doi: 10.1109/CEC.2017.7969331
50. Liu S, Zhang J, Xiang Y, Zhou W, Xiang D. A study of data pre-processing techniques for imbalanced biomedical data classification. *Int J Bioinform Res Appl*. 2020;16(3):290-318.
51. Sharma A, Rani R. An optimized framework for cancer classification using deep learning and genetic algorithm. *J Med Imaging Health Inform*. 2017;7(8):1851-1856. doi: 10.1166/jmihi.2017.2266
52. Sun KF, Sun LM, Zhou D, *et al*. XGBG: A novel method for identifying ovarian carcinoma susceptible genes based on deep learning. *Front Oncol*. 2022;12:897503. doi: 10.3389/fonc.2022.897503
53. Alharbi F, Vakanski A. Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering* (*Basel*). 2023;10(2):173. doi: 10.3390/bioengineering10020173
54. Liu R, Liu Y, Yan Y, Wang JY. Iterative deep neighborhood: A deep learning model which involves both input data points and their Neighbors. *Comput Intell Neurosci*. 2020;2020:9868017. doi: 10.1155/2020/9868017
55. Young JD, Cai C, Lu X. Unsupervised deep learning reveals prognostically relevant subtypes of glioblastoma. *BMC Bioinformatics*. 2017;18(Suppl 11):381. doi: 10.1186/s12859-017-1798-2
56. Abdullah AA, Hassan MM, Mustafa YT. A review on bayesian deep learning in healthcare: Applications and challenges. *IEEE Access*. 2022;10:36538-36562. doi: 10.1109/access.2022.3163384
57. Alakwaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res*. 2018;17(1):337-347. doi: 10.1021/acs.jproteome.7b00595
58. Sokhangouy SK, Zeinali M, Nazari E, *et al*. *Deep Learning Assisted Identification of ATP5J and ALDH1A2 Combination in RNA-Sequencing Data as a Novel Specific Potential Diagnostic Biomarker in Prostate Cancer*. PREPRINT (Version 1); 2023. Available from: https://researchsquare. doi: 10.21203/rs.3.rs-3482392/v1
59. Sur C. GSIAR: Gene-subcategory interaction-based improved deep representation learning for breast cancer subcategorical analysis using gene expression, applicable for precision medicine. *Med Biol Eng Comput*. 2019;57:2483-2515. doi: 10.1007/s11517-019-02038-2
60. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi: 10.1038/nature14539
61. Inoue M, Inoue S, Nishida T. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif Life Robot*. 2018;23(2):173-185. doi: 10.1007/s10015-017-0422-x
62. Eluri NR, Kancharla GR, Dara S, Dondeti V. Cancer data classification by quantum-inspired immune clone optimization-based optimal feature selection using gene expression data: Deep learning approach. *Data Technol Appl*. 2022;56(2):247-282. doi: 10.1108/DTA-05-2020-0109
63. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst*. 2014;27:3320-3328. doi: 10.48550/arXiv.1411.1792
64. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*. 2017;18(5):851-869. doi: 10.1093/bib/bbw068
65. Bhat RR, Viswanath V, Li X, editors. DeepCancer: Detecting Cancer Via Deep Generative Learning Through Gene Expressions. In: *2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress* (*DASC/PiCom/DataCom/CyberSciTech*). IEEE; 2017. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.152
66. Tapak L, Ghasemi MK, Afshar S, Mahjub H, Soltanian A, Khotanlou H. Identification of gene profiles related to the development of oral cancer using a deep learning technique. *BMC Med Genomics*. 2023;16(1):35.

doi: 10.1186/s12920-023-01462-6
67. Majumder S, Pal V, Yadav A, Chakrabarty A. Performance analysis of deep learning models for binary classification of cancer gene expression data. *J Healthc Eng*. 2022;2022:1122536. doi: 10.1155/2022/1122536
68. Tabares-Soto R, Orozco-Arias S, Romero-Cano V, Bucheli VS, Rodríguez-Sotelo JL, Jiménez-Varón CF. A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data. *PeerJ Comput Sci*. 2020;6:e270. doi: 10.7717/peerj-cs.270
69. Singh D, Febbo PG, Ross K, *et al*. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*. 2002;1(2):203-209. doi: 10.1016/s1535-6108(02)00030-2
70. Shah SH, Iqbal MJ, Ahmad I, Khan S, Rodrigues SJPC. *Optimized Gene Selection and Classification of Cancer from Microarray Gene Expression Data Using Deep Learning. Neural Computing and Applications*. Berlin: Springer; 2020. doi: 10.1007/s00521-020-05367-8
71. Park S, Huang E, Ahn T. Classification and functional analysis between cancer and normal tissues using explainable pathway deep learning through RNA-sequencing gene expression. *Int J Mol Sci*. 2021;22(21):11531. doi: 10.3390/ijms222111531
72. De Guia JM, Devaraj M, Leung CK, editors. DeepGx: Deep Learning Using Gene Expression for Cancer Classification. In: *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (*ASONAM*). IEEE; 2019. doi: 10.1145/3341161.3343516
73. Ibrahim M, Muhammad Q, Zamarud A, Eiman H, Fazal F. Navigating glioblastoma diagnosis and care: Transformative pathway of artificial intelligence in integrative oncology. *Cureus*. 2023;15(8):e44214. doi: 10.7759/cureus.44214
74. Chow JCL, Wong V, Li K. Generative Pre-trained transformer-empowered healthcare conversations: Current trends, challenges, and future directions in large language model-enabled medical chatbots. *BioMedInformatics*. 2024;4(1):837-852. doi: 10.3390/biomedinformatics4010047
75. Ness S, Xuan TR, Oguntibeju OO. Influence of AI: Robotics in healthcare. *Asian J Res Comput Sci*. 2024;17(5):222-237. doi: 10.9734/ajrcos/2024/v17i5451
76. Odah M. *Artificial Intelligence (AI) and Machine Learning (ML) in Diagnosing Cancer: Current Trends*. Preprints; 2024. p. 2024030433. doi: 10.20944/preprints202403.0433.v1
77. Anyanwu EC, Okongwu CC, Olorunsogo TO, Ayo-Farai O, Osasona F, Daraojimba OD. Artificial intelligence in healthcare: A review of ethical dilemmas and practical applications. *Int Med Sci Res J*. 2024;4(2):126-140. doi: 10.51594/imsrj.v4i2.755
78. Sharma R. Artificial intelligence in healthcare: A review. *Turk J Comput Math Educ* (*TURCOMAT*). 2020;11(1):1663-1667. doi: 10.61841/turcomat.v11i1.14628
79. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: A survey study from consumers' perspectives. *BMC Med Inform Decis Mak*. 2020;20(1):170. doi: 10.1186/s12911-020-01191-1
80. Vallverdú J. Challenges and controversies of generative AI in medical diagnosis. *Euphyía Rev Filos*. 2023;17(32):88-121. doi: 10.33064/32euph4957
81. Klemt C, Uzosike AC, Cohen-Levy WB, Harvey MJ, Subih MA, Kwon YM. The ability of deep learning models to identify total hip and knee arthroplasty implant design from plain radiographs. *J Am Acad Orthop Surg*. 2022;30(9):409-415. doi: 10.5435/JAAOS-D-21-00771
82. Zhang J, Li Z, Lin H, *et al*. Deep learning assisted diagnosis system: Improving the diagnostic accuracy of distal radius fractures. *Front Med* (*Lausanne*). 2023;10:1224489. doi: 10.3389/fmed.2023.1224489
83. Ando H, Chang H. A model of computing with road traffic dynamics. *Nonlinear Theory Appl IEICE*. 2021;12(2):175-180. doi: 10.1587/nolta.12.175
84. Justus D, Brennan J, Bonner S, McGough AS. Predicting the Computational Cost of Deep Learning Models. In: *2018 IEEE International Conference on Big Data* (*Big Data*). Seattle, WA, USA; 2018. doi: 10.1109/BigData.2018.8622396
85. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. *J Biomed Inform*. 2018;83:112-134. doi: 10.1016/j.jbi.2018.04.007
86. Bolhasani H, Jassbi SJ. Deep learning accelerators: A case study with MAESTRO. *J Big Data*. 2020;7(1):100. doi: 10.1186/s40537-020-00377-8
87. Che Z, Purushotham S, Khemani R, Liu Y. *Distilling knowledge from Deep Networks with applications to Healthcare Domain*. arXiv; 2015. Available from: https://arxiv.org/abs/1512.03542
88. Ali A, Ali H, Saeed A, *et al*. Blockchain-powered healthcare systems: Enhancing scalability and security with hybrid deep learning. *Sensors* (*Basel*). 2023;23(18):7740. doi: 10.3390/s23187740
89. Bebortta S, Tripathy SS, Basheer S, Chowdhary CL. DeepMist: Toward deep learning assisted mist computing framework for managing healthcare big data. *IEEE Access*. 2023;11:42485-42496. doi: 10.1109/access.2023.3266374
90. Baharani M, Biglarbegian M, Parkhideh B, Tabkhi H. Real-time deep learning at the edge for scalable reliability modeling of Si-MOSFET power electronics converters. *IEEE Internet Things J*. 2019;6(5):7375-7385. doi: 10.1109/jiot.2019.2896174
91. Nguyen AT, Drealan MW, Luu DK, *et al*. A portable, self-contained neuroprosthetic hand with deep learning-based finger control. *J Neural Eng*. 2021;18(5):056051. doi: 10.1088/1741-2552/ac2a8d
92. Hoang D, Hoang S. Deep learning - cancer genetics and application of deep learning to cancer oncology. *Vietnam J Sci*

Technol. 2022;60(6):885-928.
doi: 10.15625/2525-2518/17256

93. Lai YH, Chen WN, Hsu TC, Lin C, Tsao Y, Wu S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Sci Rep.* 2020;10(1):4679.
doi: 10.1038/s41598-020-61588-w
94. Zhu W, Xie L, Han J, Guo X. The application of deep learning in cancer prognosis prediction. *Cancers* (Basel). 2020;12(3):603.
doi: 10.3390/cancers12030603
95. Meng C, Trinh L, Xu N, Enouen J, Liu Y. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Sci Rep.* 2022;12(1):7166.
doi: 10.1038/s41598-022-11012-2
96. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246.
doi: 10.1093/bib/bbx044
97. Hage Chehade A, Abdallah N, Marion JM, Oueidat M, Chauvet P. Lung and colon cancer classification using medical imaging: A feature engineering approach. *Phys Eng Sci Med*. 2022;45(3):729-746.
doi: 10.1007/s13246-022-01139-x
98. Kiran Sree P. Deep learning for heart attack prediction. *Biomed J Sci Tech Res*. 2023;54(2):45718-45721.
doi: 10.26717/bjstr.2023.54.008522
99. Saillard C, Schmauch B, Laifa O, *et al*. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology*. 2020;72(6):2000-2013.
doi: 10.1002/hep.31207
100. Wang S, Yang DM, Rong R, *et al*. Artificial intelligence in lung cancer pathology image analysis. *Cancers* (*Basel*). 2019;11(11):1673.
doi: 10.3390/cancers11111673
101. Verma DDK. Explainable AI in healthcare: Interpretable deep learning models for disease diagnosis. *Pharma Innov*. 2019;8(3):561-565.
doi: 10.22271/tpi.2019.v8.i3j.25392
102. Wei P. Radiomics, deep learning and early diagnosis in oncology. *Emerg Top Life Sci*. 2021;5(6):829-835.
doi: 10.1042/ETLS20210218
103. Wulczyn E, Steiner DF, Xu Z, *et al*. Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLoS One*. 2020;15(6):e0233678.
doi: 10.1371/journal.pone.0233678
104. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13(1):152.
doi: 10.1186/s13073-021-00968-x
105. Hosny A, Parmar C, Coroller TP, *et al*. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS Med*. 2018;15(11):e1002711.
doi: 10.1371/journal.pmed.1002711
106. Akazawa M, Hashimoto K. *Development and Validation of Machine Learning Models for the Prediction of Overall Survival and Cancer-Specific Survival in Patients with Endometrial Cancer: An Analysis of the Surveillance, Epidemiology, and End Results (SEER) Database*. Available from: https://ssrn.com/abstract=4191367
doi: 10.2139/ssrn.4191367