Artificial Intelligence in Pathomics and Genomics of Renal Cell Carcinoma

J. Everett Knudsen, BSE, Joseph M. Rich, BS, Runzhuo Ma, MD*

KEYWORDS

• Genomics • Pathomics • Artificial intelligence • Machine learning • Kidney neoplasms

KEY POINTS

- Artificial intelligence (AI) models, using techniques such as convolutional neural networks and logistic regressions, have shown promise in renal cell carcinoma diagnosis, grading, and subtyping.
- AI has outperformed traditional methods in identifying kidney cancer biomarkers and subtype classification, particularly in clear cell renal cell carcinoma.
- Despite progress, challenges remain, including the need for consensus on best practices, computational power for large-scale models, and the creation of a ground-truth training set for model development.

INTRODUCTION

Broadly defined, artificial intelligence (AI) is the ability of a computer to model some form of human interaction. AI as a concept can be traced as far back as third-century China with the invention of a humanlike machine that seemed as if it was meant to perform simple humanlike tasks.[1] Since the 1950s, AI has progressed at a blinding speed, as human innovation in the world of computers has skyrocketed. From social media to finance to sociology, AI has left a marked and profound impact on society. In the field of urology, AI (to its broadest definition) is in practice every day as surgeons perform prostatectomies and nephrectomies with surgical robots, but AI is also finding a home in the diagnosis, grading, treatment, and survival predictability of cancer.[1] In the United States alone, there were an estimated 81,800 cases of renal cell carcinoma (RCC), resulting in approximately 15,000 deaths.[2] On identification of a renal mass, physicians need accurate, reliable methods for determining tumor subtype, grade, stage, responsiveness to pharmaceutical treatment, and the likelihood of patient survival. Luckily, clinical applications of AI have resulted in the inception of genomics and pathomics—2 fields with broad, impactful applications in the diagnosis and treatment of RCC.

Genomics and pathomics arose when machine learning (ML) algorithms were applied to gene expression patterns and pathology images, respectively. Genomics can be broadly defined as the application of ML to expressions of genes and proteins within cells of interest. Useful outcomes of genomics include specific phenotype or genotype identification, patient stratification using ML-pinpointed biomarkers, understanding gene function, and mapping the temporal biochemical significance of gene expression over time.[3] Genomics in RCC is usually leveraged at a preoperative or postoperative time point to help physicians tailor treatment options or plan for upcoming surgeries. In a similar vein, pathomics emerged when AI and ML were applied to digital pathology images. Images are more than just visual objects—they can be quantified using color scales and filtered to generate numerically

Catherine & Joseph Aresty Department of Urology, USC Institute of Urology, Center for Robotic Simulation & Education, University of Southern California, Los Angeles, CA, USA
* Corresponding author. University of Southern California Institute of Urology, 1441 Eastlake Avenue Suite 7416, Los Angeles, CA 90089.
E-mail address: marunzhuo@gmail.com

Urol Clin N Am 51 (2024) 47–62
https://doi.org/10.1016/j.ucl.2023.06.002

determinate edge patterns. Similarly to gene expression data, quantified images can be fed into ML algorithms to predict patterns of interest. Pathomics can be thought of as computers learning to recognize patterns from whole-slide pathology images and then making useful predictions after viewing novel images.[4] The following review seeks to outline the progress and applications of genomics and pathomics in RCC.

ARTIFICIAL INTELLIGENCE IN RENAL CELL CARCINOMA PATHOMICS

Pathomics is the application of ML algorithms to digital pathology slides in order to extract patterns and understand relationships that might not be readily evident to human pathologists. Over the last few decades, computer hardware and software have become faster, more efficient, and capable of storing large volumes of digital information. Applied to the world of pathology, this means pathologists can now digitize whole-slide images (WSIs) and store them for later use. When digitized, images represent a wealth of quantifiable data in the form of color values, edge detection, pixel intensity, morphology, topology, and much more. Such information can then be fed into ML algorithms to assist pathologists in diagnosing and subtyping various clinical conditions[4] (**Fig. 1**). Pathomics specific to RCC includes determining specific subtypes, assigning a Furman grade to tumors, providing cancer staging for patients, and predicting survival outcomes in patients diagnosed with RCC (=**Table 1**).

Various models have been developed that seek to shed light on the aforementioned categories. A large number of these models are considered convolutional neural networks (CNNs), which, at their core, are multilayered data processing pipelines that allow computers to extract features from images.[5] Computer vision is the most commonly used AI modeling approach in pathomics because it allows computers to detect patterns of interest from image data. The differences between CNNs lie in the number of layers available, the sharing of information between layers, and the overall efficiency of the CNN itself.[6] Classic supervised models such as Random Forests (RFs), support-vector networks (SVNs), and logistical regressions have also been developed for use in pathomics. These models are reliant on human-labeled ground-truth examples.[1] RF models improve feature selection by allowing for further building of expandable decision trees in subspaces.[7] SVNs have also been developed for WSI image classification. These are more simplistic models that map nonlinear data onto a linear feature selection space to predict a binary outcome.[8] As such, these models are limited to diagnosis between benign and malignant lesions (ie, binary) rather than subtyping or Fuhrman grading (ie, multitier). Logistical regression modeling is a simple yet highly effective technique for classifying binary outcomes that can then be applied to novel data for classification of items of interest—in this case, benign versus malignant RCC.[9] There are also multiclass regression for subtyping or staging.

Artificial Intelligence Models for Renal Cell Carcinoma Subtyping

RCC is actually a collection of kidney neoplasms that arise from different parts of the nephron.

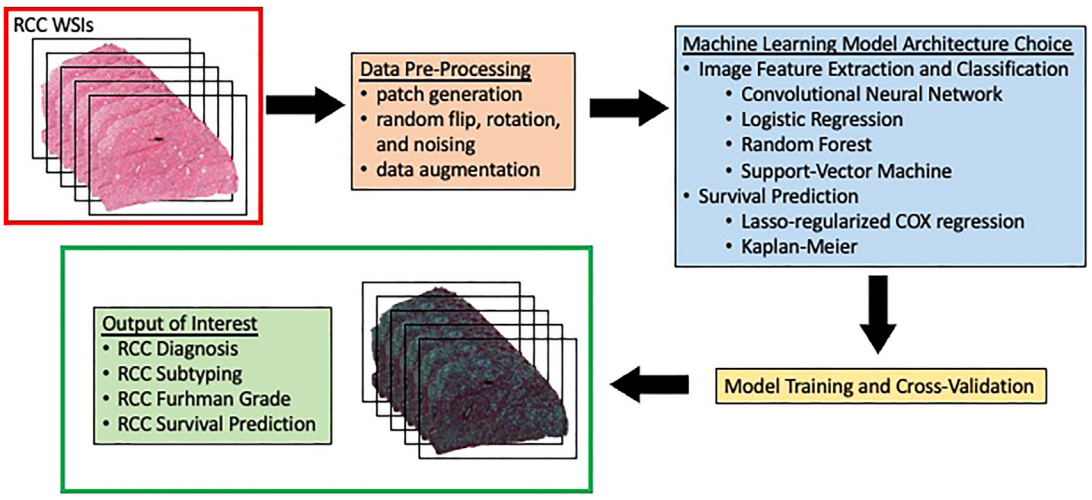

Fig. 1. The common process of using AI in pathomics.

Table 1
Summary of studies using artificial intelligence in renal cell carcinoma pathomics

Study	Application	Specific Aim(s)	Sample	Models	Performance Metrics
Azuaje et al,[79] 2019	Diagnosis	Connected H&E WSIs with protein expression using ML	WSIs: 783 (259 normal, 524 tumor); Prot: 194 (84 normal, 110 tumor)	RF, CNN, and FCNN	Accuracy Prot: 0.98 Accuracy WSIs: 095 AUC Prot: 0.99 AUC WSIs: 0.92
Chanchal et al,[22] 2023	Fuhrman Grading	Created shared residual channel (SRC)-CNN with better performance	3442 WSI patches	SRC CNN	Accuracy: 0.9014 F1 score: 0.8906
Cheng et al,[80] 2020	Diagnosis and Subtyping	Distinguished between TFE3-RCC and ccRCC WSIs using ML	148 WSIs from cases with TFE3-RCC or ccRCC	LR, SVM with linear kernel, SVM with Gaussian kernel, and RF	AUCs: LR: 0.873 RF: 0.848 SVM-L: 0.842 SVM-G: 0.894
Cheng et al,[81] 2018	Survival Prediction from Tumor Microenvironment	Examined tumor microenvironment in pRCC and predicted stage and risk index	190 TCGA samples 856 ROIs for model development	SSAE CNN	Binary outcome for 5-y survival: Stage: 0.63 Subtype: 0.66 Predicted risk index AUC: 0.78
Fenstermaker et al,[11] 2020	Diagnosis, Subtyping, and Grade	Created ML model to distinguish between normal tissue, ccRCC, pRCC, and cpRCC as well as assign Fuhrman grade	42 TCGA WSIs	CNN with 5 fully connected layers	Accuracy: Diagnostic: 0.994 Subtype: 0.975 Tumor Grade: 0.984 Diagnostic Sens.: 1.0 Diagnostic Spec.: 0.971 Diagnostic AUC: 0.98
Gondim et al,[15] 2023	Subtyping	Created a WSI patch classifier using Google AutoML Vision and deployed model on web-based API for clinical usage	252 WSIs (197 for path classifier and 55 for WSI-level tumor classification) 298,071 unique patches	Google AutoML Vision	AuPRC: 0.939

(*continued on next page*)

Table 1
(*continued*)

Study	Application	Specific Aim(s)	Sample	Models	Performance Metrics
Holdbrook et al,[77] 2018	Fuhrman Grading	Created model to classify low- vs high-grade ccRCC	59 patients; 94 training WSIs, and 20 test WSIs	AdaBoost CNN feature detection; SVM for histogram polar gradient; AdaBoost for combined	F-scores ranged from 0.78 to 0.83
Kalra et al,[26] 2020	TCGA Search Function	Created ML-based TCGA WSI search function	30,000 TCGA image database covering 25 anatomic sites and 32 cancer subtypes	Yottixel image search algorithm	Accuracy: Bladder urothelial carcinoma 93% Kidney RCC 97% Ovarian serous cystadenocarcinoma 99% Prostatic adenocarcinoma 98% Skin cutaneous melanoma 99% Thymoma 100%
Khoshdeli et al,[19] 2018	Diagnosis and Grading	Compared performance of 2 different methods of CNN to determine subtype and grade of RCC	2461 images: 796 normal, 271 fat, 42 blood, 784 stroma, 84 low-grade tumors, 484 ccRCC	GoogLeNet CNN; Shallow CNN	GoogLeNet: Precision: 0.99 Recall: 0.98 F1-score: 0.99 Shallow CNN: Precision 0.94 Recall: 0.90 F1-score: 0.92
Kruk et al,[21] 2017	Fuhrman Grading	Used a wavelet transformation preprocessing step followed by ML modeling to assign Fuhrman grade to ccRCC WSIs	94 ccRCC images	SVM with Gaussian kernel; Breiman RF	Average Sens.: 94.3% Average Spec.: 98.6%

Lu et al,[17] 2021	Subtyping	Developed a CLAM deep-learning model for subtyping RCC based on WSI-level labels. Adapted to smartphone microscopy	884 WSIs: 489 ccRCC, 284 pRCC, 111 cpRCC; BWH 135 WSIs-46 ccRCC, 46 pRCC, 43 cpRCC	CLAM deep-learning model	Average AUC: 0.991 ± 0.004 (sd.)
Ohe et al,[24] 2023	Phenotyping and Survival Prediction	Developed ALEXNET model to distinguish between clear and mixed/ eosinophilic ccRCC and predict survival rate.	TCGA 435 WSIs, 95 WSIs for validation	Deep CNN called ALEXNET	Average AUC: 0.929 (95% CI 0.88–0.98) Average survival rate: Mixed/eosinophilic 54.3% Clear 80.9%
Tabibu et al,[13] 2019	Diagnosis, Subtyping, and Survival Prediction	Developed CNN ML model for RCC diagnosis and subtyping. Correlated these results for HR prediction	1027 ccRCC, 303 pRCC, 254 cpRCC, 379, 47, 83 normal slides from each subtype	CNN with DAG-SVM on the fully connected CNN layer; Lasso-regularized Cox for survival	Accuracy: ccRCC vs benign: 93.39% cpRCC vs benign: 87.34% Average accuracy for all subtyping 94.07% Lasso-Cox HR: 2.265 (95% CI 1.5343–3.343)
Tian el al,[20] 2019	RCC Grading and Survival Prediction	Developed Lasso-regularized Cox model to assign Fuhrman grade to ccRCC and predict HR	305 ccRCC WSIs from TCGA	Lasso regression model using linear regression with L1 regularization; Cox proportional hazard model for survival prediction	Accuracy: 83.3% grade prediction Sens.: 84.6% grade prediction Spec.: 81.3% grade prediction Avg. AUC: 0.84 HR: 2.05 (95% CI 1.21–3.47)

Abbreviations: CI, confidence interval; H&E, hematoxylin and eosin; HR, hazard ratio; SSAE, stacked sparse autoencoder.

Each subtype of RCC has distinct patterns of gene expression and histologic features. Of the subtypes of RCC, the 3 most common subtypes are clear cell RCC (ccRCC), papillary RCC (pRCC), and chromophobe RCC (cpRCC).[10] Besides the RCC subtypes, oncocytoma, metanephric adenoma, and fat-poor angiomyolipoma are benign conditions that present similarly to RCC. As such, physicians are interested in distinguishing between benign and malignant neoplasms and between RCC subtypes using histologic images. Research in pathomics has provided a variety of models targeted toward this very task.

Diagnosis of kidney cancer starts with a patient's presenting symptoms and subsequent medical imaging to look for kidney masses. After mass detection, neoplasms may be biopsied by a needle or, when appropriate, removed entirely with full or partial nephrectomy.[10] WSIs are then generated from biopsy or resected tumor samples. In the world of pathomics, Fenstermaker and colleagues created a CNN capable of distinguishing between normal renal parenchyma and malignant tissue. Data were obtained from 42 patients who had samples stored in the National Institutes of Health's The Cancer Genome Atlas Data Access Portal (TCGA), a WSI database used by many cancer researchers.[11,12] Tabibu and colleagues also created a CNN with support vector machine (SVM) layer to classify normal versus malignant kidney tissue.[13] In 2021, Zhu and colleagues created a CNN capable of classifying normal kidney parenchyma, oncocytoma, and RCC.[14] Gondim and colleagues then produced a model in Google AutoML Vision that could also distinguish metanephric adenoma from the aforementioned kidney tissue types.[15]

Besides distinguishing between benign and malignant neoplasms, subtyping of RCC has been a great focus in RCC pathomics. The previously mentioned CNNs created by Tabib and colleagues and Fenstermaker and colleagues were also capable of distinguishing between ccRCC, pRCC, and cpRCC.[13] Ghaffari Laleh and colleagues and Gondim and colleagues also produced models with similar subtyping capabilities.[15,16] Focusing on an extremely rare RCC subtype, Cheng and colleagues used logistic regression, SVM, and RF to distinguish between TFE3 Xp11.2 translocation RCC (TFE3-RCC) and ccRCC.

Finally, Lu and colleagues have developed a model that also uses deep learning but requires far less time by trained pathologists to label the training dataset. Using a method called clustering-constrained-attention multiple-instance learning (CLAM), WSIs are labeled at the image level rather than in smaller patches. These WSI labels then allow the model to look for regions of interest, which caused the pathologist to assign the label in the first place. Therefore, CLAM approaches reduce the time necessary to label slides in a spatial format and reduce noise in the training dataset. The model by Lu and colleagues was capable of binary normal versus malignant predictions and subtyping classification predictions and outperformed classic weakly supervised approaches.[17]

Artificial Intelligence Models for Renal Cell Carcinoma Fuhrman Grading

When diagnosed with RCC both patients and physicians are also concerned with grading of the tumor, which gives an idea of the aggressiveness associated with the specific neoplasm. The most widely used grading scale is the Fuhrman system, which assesses nuclear morphology as a prognostic indicator for RCC.[10] The Fuhrman scale assigns 4 grades, with a higher grade indicating a worse prognosis. Similar to RCC subtyping, Fuhrman grading models use a variety of model types including CNNs, logistic regressions, RFs, SVMs, multiple-instance learning models, and CLAM models. These models are, again, considered computer vision approaches because they allow the computer to "see" nuclear morphology and assess severity.

In 2014, Yeh and colleagues used a Kernel regression model to determine nuclear size variations and then correlate these size variations with Fuhrman grades. The model was able to distinguish between low- (Fuhrman grades 1 and 2) and high-grade (Fuhrman grades 3 and 4) with high accuracy.[18] Khoshdeli and colleagues created a shallow CNN and a deep CNN based on GoogLeNet to distinguish between normal tissue, low-, and high-grade ccRCC. They found that both models were capable of making such distinctions, but the GoogLeNet-based model outperformed the shallow CNN.[19] Tian and colleagues presented a Lasso model that was capable of predicting a 2-tiered Fuhrman grade for ccRCC based on 26 model features.[20] In 2017, Kruk and colleagues presented a model capable of assigning Fuhrman grades 1 to 4 rather than the high- and low-grade distinctions seen in previous studies. This model used wavelet transformation in a preprocessing step to reduce noise and improve edge detection. SVM and RF classifiers were then applied to the preprocessed WSIs to assign a Fuhrman grade to WSIs. High accuracy for both SVM and RF classifiers was achieved using only 11 model features.[21]

Fenstermaker and colleagues also presented a 4-tiered Fuhrman grading model in which they used a CNN with a learned first layer, a pixel compressor, and several further layers with full sharing of information.[11] Further, Chanchal and colleagues have also developed the Renal Cell Carcinoma Grading Network (RCCGNet), which is a CNN with a shared residual block that was trained on 722 RCC WSIs. The model assigns a 5-tiered grade in that it can distinguish between normal tissue as well as between Fuhrman grades 1 and 4.[22] Each advancement made by researchers in pathomics represents further advancement toward AI helping pathologists understand the severity of patients' disease with high confidence.

Artificial Intelligence Models for Renal Cell Carcinoma Survival Prediction

For patients with any form of neoplastic disease, prognosis is arguably the most important metric that they would like to understand. Survival prediction is, perhaps, the thing that is of utmost clinical relevance when it comes to pathomics and genomics. In RCC pathomics, several models have been developed that seek to assign a quantifiable value to survival likelihood with the hope that physicians can use these predictions to provide better counseling and support to their patients. For researchers in pathomics, development of a survival prediction algorithm is reliant on the collection of WSIs taken from resected or biopsied tumor specimens and on patient factors such as age, biological sex, treatment course, and postoperative survival outcomes. WSI and patient outcome information are then correlated to develop a model capable of predicting survival after viewing novel pathology images.[20]

As previously discussed, Tian and colleagues developed a Lasso model to predict and assign RCC Fuhrman grades from WSIs. The 160 WSI samples (42 training, 116 validation) used in this study were pulled from TCGA, which also includes deidentified patient information important for survival predictability. Tian and colleagues then constructed crude and adjusted Cox proportional hazard models and validated them using the additional 116 training WSIs. The Cox models were capable of predicting an overall survival percentage.[20] In a similar vein, Tabibu and colleagues also predicted survival using a similar approach as Tian and colleagues. Tabibu and colleagues created a CNN capable of determining RCC subtype, and this CNN detected several features of interest that were also useful for predicting survival. They calculated the risk index of each patient using lasso-regularized Cox modeling for each feature used in the subtyping classifier. Tumor shape and nuclei shape features were significantly associated with patient survival. Chen and colleagues took a different approach in which they used a digital pathology software called QuPath to select a variety of features based on cell morphology and create a ML-based pathomics signature (MLPS) for each WSI analyzed. They then used Cox survival regression analysis to predict disease-free survival[23]; this is the first instance of MLPS generation used to predict survival in a patient population. Ohe and colleagues also developed a survival prediction algorithm targeted toward predicting prognosis in clear versus eosinophilic subtypes of ccRCC. They created a deep CNN from 435 TCGA ccRCC WSIs that distinguished between clear and eosinophilic phenotypes in ccRCC and assigned an AI score to each phenotype. Kaplan-Meier survival analysis was able to predict worse survival rates in patients with mixed/eosinophilic-predominant subtypes versus subtypes. Patients with higher AI scores as determined by the deep CNN had worse survival prognosis, which was validated by real-world survival outcomes.[24] A final survival prediction model was developed by Cheng and colleagues. For this model, researchers focused on classifying the topographic features of the tumor microenvironment in pRCC. Historically, pRCC is not as studied as more common subtypes of RCC such as ccRCC, so this group hoped to shed more light on how the tissue surrounding the tumor affects prognosis. They used an unsupervised approach with a neural network called a stacked sparse autoencoder for feature extraction, then used K-means clustering and Delaunay triangulation for the identification of nuclei morphology patterns. Features from WSI images were then used to build a lasso-regularized Cox regression model to predict risk indices for patients. Certain tumor microenvironment topologies increased risk in pRCC.[25]

Clinical Applications of Artificial Intelligence in Renal Cell Carcinoma Pathomics

As discussed earlier, RCC pathomics has given physicians a wide variety of models capable of determining RCC subtype, grade, and survival prediction. However, it is essential that pathomics moves from theoretic exercises to real-world, clinical applications in order to be useful when it comes to treating patients. Of the models outlined earlier, a few have provided the first steps toward clinical applications beyond just the development of AI models. In addition to improved efficiency

and applicability to binary and subtyping tasks, Lu and colleagues' CLAM model generated WSI heatmap overlays that were able to show pathologists the regions of interest that led the model to assign a specific subtype. These heatmaps are highly clinically relevant because they allow pathologists to quickly target areas of a slide that show abnormal cellular architecture or nuclear atypia, meaning pathologists' cognitive loads are reduced.[17] Lu and colleagues were also able to feed their model WSIs taken by smartphone cameras, with the model producing similar results to those taken by typical light microscopy. Applications such as these might be deployed in resource-limited areas where either expert pathologists or highly technical equipment is not readily available. Lu and colleagues imagine a scenario in which a physician in a rural area could submit a digitized biopsy specimen to an online application programming interface (API) and receive a subtype in a matter of seconds rather than sending it to be read by pathologist.[17] Web-hosted APIs for pathologic diagnostic aid could help pathologists efficiently and effectively assign diagnoses in the future.

Another clinical application that extends beyond RCC is a model developed by Kalra and colleagues that is capable of searching all the WSIs uploaded to TCGA. By building a WSI-based search function, pathologists might be able to pull WSIs from uploaded patient samples and compare them with a novel sample they are working with. In this way, they might be able to view other examples of specific subtypes or morphologies before assigning a diagnosis to a new patient case.[26]

ARTIFICIAL INTELLIGENCE IN RENAL CELL CARCINOMA GENOMICS

AI models have been used in a variety of genomic problems including gene expression analysis, transcription factor binding site identification, exon splicing patterns, finding disease-causing genetic variants, and predicting chromatin structure, among a wide array of additional applications.[27–31] Genomics lends itself nicely to ML approaches due to the availability of patterns in genomics data, large dataset sizes, and ability to combine model results with prior knowledge and/or experimental validation. Historically, linear models have had success in supervised (eg, regression, SVMs, RF) and unsupervised (eg, K-means, principal component analysis (PCA), t-distributed stochastic neighbor embedding [t-SNE]) learning tasks in bioinformatics due to their simplicity and robustness.[32–36] More recently, with increasing advances in both biological data and computing power, deep learning has been gaining traction in this field as a means of uncovering complex genomic relationships.

Artificial Intelligence Models for Genomics

Some of the most popular deep-learning models for supervised learning in genomics include fully connected neural networks, convolutional neural networks, recurrent neural networks, and graph convolutional networks[3] (**Fig. 2**). Each architecture has pros and cons involving computational cost, invariance, interpretability, and prediction quality.

Fully connected networks are the standard feedforward neural network in which all nodes of each layer are connected to all nodes of each adjacent layer. In the context of genomics, the input for a neural network is the one-hot encoding of a DNA sequence. Nodes take as input the weighted average of all nodes from the previous layer, and the result is passed through a (likely) nonlinear activation function before the process is repeated for the following layer. The size of the final layer of the network corresponds to the value or number of classes being predicted. These networks have been used in a wide range of genomic applications including gene expression, splicing patterns, and sequence analysis.[27–29]

Convolutional neural networks pass small filters (ie, low-dimensional matrices) with shared parameters as sliding windows over input data, allowing for translation invariance (ie, relative position of a portion of data does not alter computation with the filter). The dot product between a portion of input data and the filter is performed, and then the filter slides to the next section of input data. Each filter represents a small pattern, which in genomics possesses input data comprised of genetic sequences, which could identify specific motifs. One of the main benefits of this sliding window approach includes saving on computational costs as a result of parameter sharing. Because these filters possess relatively understandable qualities to human interpretation, CNNs also possess a degree of explainability by visualizing the output of a filter, finding which sequence maximally activates a filter, or nullifying a filter and measuring the impact on the model's predictions.[37–39] The semantic value possessed by these filters allows domain knowledge to affect filter design, including cases where certain filters are intentionally initialized in order to seek out specific motifs. Some applications of CNNs include predicting transcription factor binding sites, DNA methylation states, microRNA (miRNA) targets, and pharmacogenomic properties.[6,40–43]

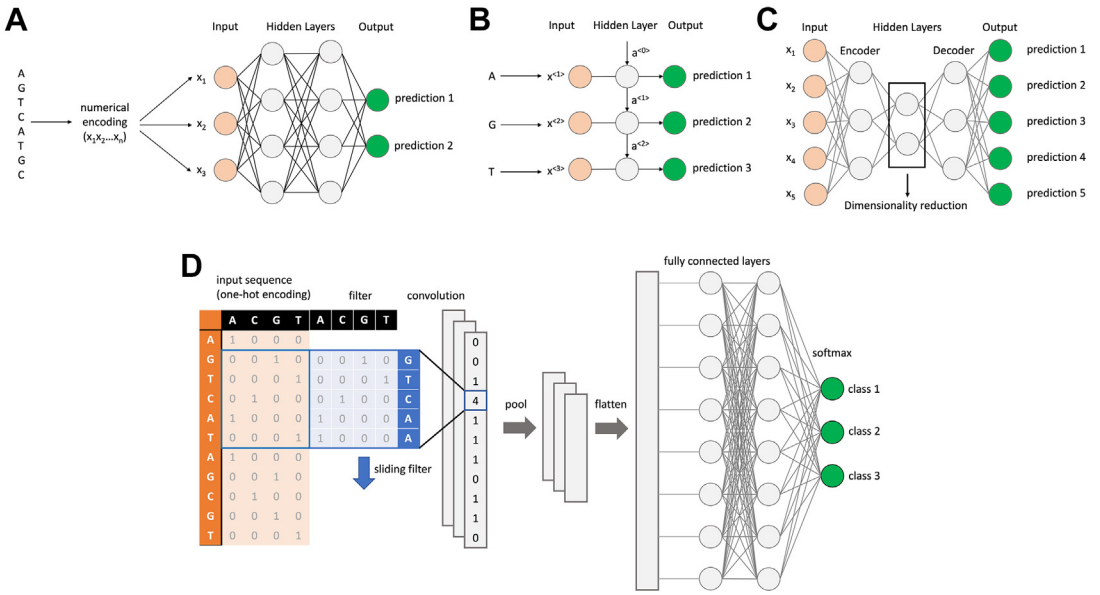

Fig. 2. Popular neural network architectures in genomics. (*A*) Fully connected feedforward neural network shows a genomic sequence as input data and binary classification output. Each input sequence is represented as a string of numbers through numerical encoding (eg, one-hot, binary, ordinal, learned embeddings), where each number serves as an input feature (x_1-x_n, where n is the number of features). Only the first 3 input features are shown. (*B*) Recurrent neural network shows individual nucleotides of a sequence as input data ($x^{<1>}$-$x^{<3>}$), with calculation of subsequent hidden layer (*N*) based on both input at that layer ($x^{<n>}$) and activation from the previous layer ($a^{<n-1>}$). (*C*) Autoencoder demonstrates encoding and decoding layers with potential for dimensionality reduction. (*D*) Convolutional neural network shows input data as (one-hot encoded) sequence with convolution operation, pooling layers, data flattening, and fully connected layers resulting in multiclass classification output. All figures show input layer (*orange*), hidden layers (*gray*), and output layer (*green*).

Recurrent neural networks pass information from temporally earlier information as input to the network trained for the subsequent data input value, thus allowing the network to possess memory. This memory property allows the network to retain time invariance (ie, is not the absolute position of an input sequence that determines prediction, but rather the context of the sequence in which the input appears). Some applications of RNNs include single-cell DNA methylation states,[44] retinol-binding protein binding,[45] and DNA accessibility.[46,47]

Graph convolutional networks apply deep learning to graphical input data. These networks apply convolution operations and subsequent nonlinear activation functions to neighboring groups of nodes to predict behaviors relating to interactions. Applications include protein-protein interaction modeling and gene expression network analysis.[48–50]

A popular deep-learning model for unsupervised learning includes autoencoders. These models use the same data as both input and output, with a hidden bottleneck layer that forces reproduction of input data with removal of redundant features. Autoencoders can be applied to impute missing data and for dimensionality reduction among other uses.[51–53]

To address problems of small or homogeneous datasets, 2 popular data augmentation techniques include transfer learning and generative models. Transfer learning involves pretraining a model with another dataset to provide initial parameter values, which may improve model predictions compared with random initialization.[54] Generally, the larger the transferred dataset and the more similar to the user's dataset, the better the predicted performance boost. Transfer learning has shown promise in biological image classification and sequence-based prediction of chromatin accessibility.[55,56] Generative models include variational autoencoders (VAEs) and generative adversarial networks (GANs). VAEs are autoencoders with probabilistic encoding, allowing for variations in decoding and data generation. VAEs have had success in RNA-seq analysis and predicting drug response.[57,58] GANs are designed adversarially, in which a generator tries to "trick" a discriminator into confusing generated data as being real. GANs have been used in sequence generation and scRNA-seq simulation.[59–61]

Model interpretability remains an ongoing challenge in deep learning due to the obscure meaning of each parameter's influence on model predictions. Genomics provides no exception to the difficulty of the interpretability problem, yet it underscores the problem's importance due to explainability of predictions often being essential to uncover biological meaning in relationships or pathways. Numerous approaches to interpretability have been explored in genomics. In addition to CNN-specific node-based strategies enumerated earlier, some architecture-agnostic methods for tackling interpretability include implementation of attention mechanisms to focus on regions of particular interest based on prior knowledge, in-silico mutagenesis (observing degree of change in model prediction based on induced point mutation in the data), motif-embedding (observing degree of change in model prediction after artificially embedding a motif in an unnatural location), gradient magnitude calculation (where higher magnitude indicates high degree of importance of that region), and feature interaction identification through simulated mutagenesis.[39]

In RCC in particular, AI has been used in a variety of ways. PCA has been used to cluster RCC by immune subtypes and weighted correlation network analysis with RF to predict gene expression of distinguishing hub genes.[62] Four supervised learning algorithms—J48, RF, SMO, and Naïve Bayes—could discern early versus late-stage ccRCC based on transcriptomic signature, and feature selection predicted 62 of the most important distinguishing genes between the 2 groups.[63] t-SNE analysis revealed a mitochondrial genetic signature that spanned histopathologic subgroups and predicted worse survival for patients with RCC.[64] The Min-Redundancy and Max-Relevance algorithm was used to pick a profile of 13 genes highly correlated with RCC patient outcomes.[65] Other combinations of supervised and unsupervised learning help predict unique genetic subtypes and gene signatures of patients with RCC, with the TCGA-KIRC population as a popular dataset for input.

Artificial Intelligence in Identifying Renal Cell Carcinoma Biomarkers

The application of AI has shown promising results, particularly in biomarker identification and subtype classification. Recent studies have proposed novel methods that leverage AI for kidney cancer biomarker identification and subtype classification, demonstrating superior performance compared with traditional methods.

Liu and colleagues presented a study on the use of bioinformatics tools and a neural network model for identifying biomarkers in ccRCC.[66] The investigators used a 2-step approach. First, they identified differentially expressed genes (DEGs) between ccRCC and normal renal tissues using data from the Gene Expression Omnibus (GEO) database. They then constructed a protein-protein interaction network of the DEGs and screened for hub genes using cytoHubba. Ten hub genes were identified, including AURKB, CCNA2, TPX2, and NCAPG, which were found to be highly expressed in ccRCC compared with normal renal tissue. In the second step, the investigators used a neural network model to verify the relationship among these genes. The model was trained using data from 10 ccRCC tumor samples and 10 normal kidney tissues. The results from the neural network model showed strong correlations between the hub genes, validating their potential as biomarkers for ccRCC. The study demonstrated the potential of AI in enhancing the identification and verification of biomarkers in ccRCC.

AI was used to analyze DNA methylation and gene expression data as well. Malouf and colleagues used a supervised clustering approach to analyze promoter DNA methylation and gene expression using a 56 genes epi-signature in TCGA dataset of ccRCC and chromophobe samples.[67] This approach allowed them to identify distinct clusters based on the methylation patterns and gene expression profiles of the samples. Furthermore, unsupervised clustering was used to analyze DNA methylation using CpG sites located in promoter CpG islands and outside promoter CpG islands. This analysis revealed 2 epi-clusters with distinct characteristics. One cluster (C1) contained almost all tumors with benign potential, whereas the other cluster (C2) contained tumors with potential malignant behavior.

Pirmoradi and colleagues used miRNA data, which are often high-dimensional and contains many irrelevant and redundant features.[68] To address this, the researchers used a filter method for feature selection, specifically the Arithmetical Mean Geometric Mean measure. This method, with its low computational cost, effectively identifies the most discriminant miRNAs as significant features, thereby enhancing the performance of disease or subtype classification. Then the study introduces a self-organizing deep neuro-fuzzy system for the classification task. This system is designed to overcome common challenges in the field, such as the curse of dimensionality, low sample numbers, and unbalanced data. Remarkably, the proposed classifier achieved an average classification accuracy of 93.2%, sensitivity of 92.4%, and specificity of 98.1% in test data. These results indicate the system's ability to classify kidney

cancer subtypes with high accuracy based on complex rules obtained using deep-learning algorithms. Another group applied the tensor decomposition (TD)-based unsupervised feature extraction method to analyze messenger RNA (mRNA) and miRNA expression profiles.[69] The AI-driven analysis identified miRNA signatures and their associated genes, which were found to be involved in cancer-related pathways. Moreover, 23 genes were significantly correlated with the survival of patients with ccRCC. The investigators demonstrated that the results are robust and does not highly depend on the databases selected. Compared with traditional supervised methods tested, TD achieves much better performance in selecting prognostic miRNAs and mRNAs; this suggests that integrated analysis using the TD-based unsupervised feature extraction technique is an effective strategy for identifying prognostic signatures in cancer studies.

Artificial Intelligence and Renal Cell Carcinoma Liquid Biopsy

Liquid biopsy is a noninvasive diagnostic tool that is gaining traction in the management of RCC. It involves the analysis of circulating tumor cells, circulating tumor DNA, and other biomarkers present in body fluids, such as blood. Liquid biopsy offers a real-time snapshot of the tumor's genetic landscape, allowing for early detection, monitoring disease progression, and assessing treatment response. Combining with AI, it holds significant promise for personalized medicine in RCC, enabling clinicians to tailor treatment strategies based on the unique molecular profile of each patient's tumor.

Iwamura and colleagues presented a study on the use of AI for diagnosing urologic diseases, including RCC, based on blood sample.[70] The study analyzed immunoglobulin N-glycan signatures from 100 serum subjects with RCC. The data were used in a supervised ML model to establish a scoring system that gave the probability of the presence of RCC. The results indicated that the RCC score could be a promising biomarker for the early diagnosis of RCC and for differentiating between invasive renal pelvis cancer and RCC. The score showed excellent diagnostic accuracy at any pathologic stage. However, further external validation trials are needed to validate the urologic disease-specific scoring system in routine clinical practice.

Manzi and colleagues used AI techniques in conjunction with lipidomics for the early detection of ccRCC.[71] The researchers developed an ML model that was trained to identify patterns in the lipid profiles of patients, which could potentially indicate the presence of ccRCC. The model was tested and validated using an independent set of patient data, demonstrating promising results in terms of accuracy, sensitivity, and specificity. The investigators also attempted to apply this model to other types of RCC, although the sample size for these types was limited. The results of this study suggest that this ML approach could be a valuable tool for early ccRCC diagnosis, pending further validation in larger and more diverse patient cohorts.

Artificial Intelligence and Renal Cell Carcinoma Staging

Bhalla and colleagues used AI techniques to classify early and late-stage patients of ccRCC based on gene expression data.[72] The study used ML for feature selection, reducing the number of features from 19,166 to 64 using a software package called Weka. They also developed prediction models using SVM and RFs and evaluated their performance using 10-fold cross-validation. The researchers also presented a Web platform called CancerCSP, where users could provide gene expression data and predict whether the cancer was in the early or late stage. This application of ML provided better insights to understand the mechanisms responsible for metastasis in various cancers.

Artificial Intelligence and Renal Cell Carcinoma Survival Prediction

Recent studies have demonstrated the use of AI in developing prognostic models for RCC. These models leverage large-scale genomic data to identify differentially expressed genes and signatures that can predict disease progression and patient survival.

Chen and colleagues used ML techniques to develop a prognostic model for patients ccRCC.[73] The researchers identified 333 DEGs between ccRCC and normal tissues from the GEO database. They used univariate Cox regression analysis to retrieve the survival-related DEGs and further used multivariate Cox regression with the LASSO penalty to identify potential prognostic genes. A 7-gene signature was identified, including APOLD1, C9orf66, G6PC, PPP1R1A, CNN1G, TIMP1, and TUBB2B. The seven-gene signature was evaluated in the training set, internal testing set, and external validation using data from the ICGC database. The Kaplan-Meier analysis showed that the high-risk group had a significantly shorter overall survival time than the low-risk group in the training, testing, and ICGC datasets. The researchers concluded that the 7-gene

signature can serve as an independent biomarker for predicting prognosis in patients with ccRCC.

Singh and colleagues have developed a ML model to predict the progression of pRCC from early to late stages using RNA sequencing data.[74] The team used a ML pipeline incorporating different feature selection algorithms and classification models. They identified 80 genes that are consistently altered between stages by different feature selection algorithms, which are related to cellular components such as centromere, kinetochore and spindle, and biological process mitotic cell cycle. The AI model developed in this study demonstrates the potential of ML in providing valuable insights into the progression of pRCC.

THE INTERACTION BETWEEN PATHOMICS AND GENOMICS AND THE EMERGENCE OF MULTIOMICS

The integration of AI with multiomics data is revolutionizing the field of RCC. Recent studies have demonstrated the potential of ML and deep-learning techniques in extracting meaningful insights from complex biological data, including genomics, proteomics, and pathomics.

Singh and colleagues used AI techniques to investigate the methylation patterns of pRCC and their relationship with gene expression using multiomics data.[75] ML models were used to analyze gene expression (RNA), DNA methylation, and clinical information. AI techniques facilitated the analysis of various representations of methylation data and enabled functional enrichment analysis to uncover biological processes associated with PRCC. The findings highlight the potential of AI in enhancing our understanding of pRCC by integrating multiomics data.

Ning and colleagues focused on improving the prediction of prognosis in ccRCC using multiomics data.[76] They sought to address the limitations of traditional methods that relied on hand-crafted features and single-modal data. Drawing inspiration from the success of CNNs in medical image analysis, they proposed a novel framework that combines deep features extracted from computed tomography/histopathological images with eigengenes derived from functional genomic data. This approach outperformed models based on single-modality features, effectively stratifying patients into high- and low-risk subgroups. The study also explored the relationship between deep image features and eigengenes, offering insights into the interpretation of deep image features using genomic data.

Holdbrook and colleagues presented an automated image-based system that leveraged AI to classify ccRCC slides by quantifying nuclear pleomorphic patterns.[77] The system, which quantified nuclear patterns, was tested on tissue slides from 59 patients, with results correlating to a multigene assay-based scoring system. The AI used a "Fraction Value" (FV) score, with a high correlation found between the FV predicted and the multigene score.

Ing and colleagues developed an ML approach to identify latent vascular phenotypes that could predict the outcome of renal cancer.[78] They used a 2-step framework for quantitative imaging of tumor vasculature to derive a spatially informed, prognostic gene signature. The algorithms they developed classified endothelial cells and generated a vascular area mask in hematoxylin and eosin micrographs of ccRCC cases from TCGA. The investigators successfully applied digital image analysis and targeted ML to develop prognostic, morphology-based, gene expression signatures from the vascular architecture. This novel morphogenetic approach has the potential to improve previous methods for biomarker development.

Azuaje and colleagues used a deep-learning CNN to find prognostic correlations between patterns of protein expression and histopathology images.[79] They found that certain proteomics patterns cause visually appreciable changes in pathology slide findings; this is another step toward connecting gene and protein expression patterns with pathomics modeling and will help pathologists make faster, more robust diagnoses in the future.

These studies underscore the transformative potential of AI in RCC research, particularly in the realm of multiomics data integration and prognosis prediction. The application of AI in this context could significantly enhance diagnostic accuracy, enable personalized treatment strategies, and ultimately improve patient outcomes.

CHALLENGES AND FUTURE DIRECTION

In this narrative review, the authors have discussed a wide array of AI models and their various applications to problems in RCC. Both pathomics and genomics models can help diagnose malignant disease and predict survival in RCC, but there are still a few challenges to overcome. As demonstrated by the number of models outlined in this review, there are a wide variety of computational approaches to consider when undergoing model development. However, there is no consensus on the best approach—some models may offer slight performance edges over others, but these enhancements are not something that patients

coming into the urology clinic will necessarily understand. To bridge this gap, researchers must make sure to ground the theory behind their work in practical clinical applications. The first steps are being made in the right direction, but we encourage researchers and physicians to work together to understand how theoretic models can be deployed in real-world scenarios to make greater impacts on patient care. Secondly, projects in pathomics and genomics also require an enormous amount of computing power to predict outcomes of interest. Institutions and hospital groups are often resource-limited and might lack the computational power necessary to deploy and maintain large-scale clinical models. Therefore, computational needs might be met by the use of commercially available servers such as Google AutoML Vision, which allows for the web-based development and deployment of clinical AI models.[15] Another limitation of AI modeling generally is the need to construct a ground-truth, human-verified training set for model development. For pathomics and genomics, this is a time- and labor-intensive task, especially as more and more tumor WSIs are generated and novel genes are discovered. In pathomics, approaches in WSI labeling with CLAM computing lighten the workload, but ground-truth training set development still represents a bottleneck.[17] Finally, the performance of AI models highly depends on the training data, which may result in unexpected bias. For example, in genomics, race may limit the generalizability of a fully trained AI model when applied to patients coming from a minority racial group. More diverse datasets can mitigate bias in this regard.

SUMMARY

The intersection of AI models with histopathology images and gene expression patterns has produced the rapidly expanding fields of pathomics and genomics, respectively. Applications of pathomics and genomics in RCC have given researchers and clinicians new tools for diagnosis and subtyping of kidney tumors. It has also allowed for the development of more robust survival prediction models that, in the future, can help patients and their families better understand the prognosis of a new RCC diagnosis. ML has also helped uncover new gene expression patterns specific to different subtypes and grades of RCC. These models are helping researchers better understand the biological origins of RCC as well as uncover potential avenues for treatment, particularly targeted medical therapies. Pathomics and genomics are also being used in combination, thanks to the ability of AI model to deal with multimodal data. With AI, patients and physicians should look forward to the future discoveries and innovations to come in the growing fields of pathomics and genomics, to shed light in the field of RCC.

CLINICS CARE POINTS

- AI has been used to assist in pathological renal cell carcinoma cancer diagnosis, subtyping, Fuhrman grading, and survival prediction in the research setting.
- A synergetic integration of AI and genetics assists basic science research in identifying renal cell carcinoma biomarkers, classifying genetic subtypes, and enabling multiomics. This powerful integration also holds promise in clinical application such as liquid biopsy.
- While AI carries substantial potential to support clinicians with the pathologic diagnosis and genomic classification of renal cell carcinoma, it is imperative to proceed with rigorous clinical trials and obtain FDA approval.

DISCLOSURES

The authors declare no commercial or financial conflicts of interest or any funding sources.

FUNDING/SUPPORT

No funding.

REFERENCES

1. Hamet P, Tremblay J. Artificial intelligence in medicine. Metab Clin Exp 2017;69:S36–40.
2. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. CA Cancer J Clin 2023;73(1):17–48.
3. Eraslan G, Avsec Ž, Gagneur J, et al. Deep learning: new computational modelling techniques for genomics. Nat Rev Genet 2019;20(7):389–403.
4. Niazi MKK, Parwani AV, Gurcan M. Digital Pathology and Artificial Intelligence. Lancet Oncol 2019;20(5): e253–61.
5. Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. Pattern Recognit 2018;77:354–77.
6. Li Z, Liu F, Yang W, et al. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Trans Neural Netw Learn Syst 2022; 33(12):6999–7019.
7. Ho TK. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition 1995;1:278–82.

8. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
9. Subasi A. Machine learning techniques. Academic Press; 2020. p. 91–202.
10. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. Lancet 2009;373(9669):1119–32.
11. Fenstermaker M, Tomlins SA, Singh K, et al. Development and Validation of a Deep-learning Model to Assist With Renal Cell Carcinoma Histopathologic Interpretation. Urology 2020;144:152–7.
12. Grossman RL, Heath AP, Ferretti V, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med 2016;375(12):1109–12.
13. Tabibu S, Vinod PK, Jawahar CV. Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning. Sci Rep 2019;9(1):10509.
14. Zhu M, Ren B, Richards R, et al. Development and evaluation of a deep neural network for histologic classification of renal cell carcinoma on biopsy and surgical resection slides. Sci Rep 2021;11(1):7080.
15. Gondim DD, Al-Obaidy KI, Idrees MT, et al. Artificial intelligence-based multi-class histopathologic classification of kidney neoplasms. J Pathol Inform 2023;14:100299.
16. Ghaffari Laleh N, Muti HS, Loeffler CML, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. Med Image Anal 2022;79:102474.
17. Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021;5(6):555–70.
18. Yeh FC, Parwani AV, Pantanowitz L, et al. Automated grading of renal cell carcinoma using whole slide imaging. J Pathol Inform 2014;5(1):23.
19. Khoshdeli M, Borowsky A, Parvin B. Deep Learning Models Differentiate Tumor Grades from H&E Stained Histology Sections. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society. EMBC); 2018. p. 620–3.
20. Tian K, Rubadue CA, Lin DI, et al. Automated clear cell renal carcinoma grade classification with prognostic significance. PLoS One 2019;14(10):e0222641.
21. Kruk M, Kurek J, Osowski S, et al. Ensemble of classifiers and wavelet transformation for improved recognition of Fuhrman grading in clear-cell renal carcinoma. Biocybern Biomed Eng 2017;37(3):357–64.
22. Chanchal AK, Lal S, Kumar R, et al. A novel dataset and efficient deep learning framework for automated grading of renal cell carcinoma from kidney histopathology images. Sci Rep 2023;13(1):5728.
23. Chen S, Jiang L, Gao F, et al. Machine learning-based pathomics signature could act as a novel prognostic marker for patients with clear cell renal cell carcinoma. Br J Cancer 2022;126(5):771–7.
24. Ohe C, Yoshida T, Amin MB, et al. Deep learning-based predictions of clear and eosinophilic phenotypes in clear cell renal cell carcinoma. Hum Pathol 2023;131:68–78.
25. Cheng J, Mo X, Wang X, et al. Identification of topological features in renal tumor microenvironment associated with patient survival. Bioinforma Oxf Engl 2018;34(6):1024–30.
26. Kalra S, Tizhoosh HR, Shah S, et al. Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. NPJ Digit Med 2020;3:31.
27. Xiong HY, Alipanahi B, Lee LJ, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 2015;347(6218):1254806.
28. Jha A, Gazzara MR, Barash Y. Integrative deep models for alternative splicing. Bioinforma Oxf Engl 2017;33(14):i274–82.
29. Ezziane Z. Applications of artificial intelligence in bioinformatics: A review. Expert Syst Appl 2006;30(1):2–10.
30. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinforma Oxf Engl 2015;31(5):761–3.
31. Chen C, Hou J, Shi X, et al. DeepGRN: prediction of transcription factor binding site across cell-types using attention-based deep neural networks. BMC Bioinf 2021;22(1):38.
32. Liu S, Lu M, Li H, et al. Prediction of Gene Expression Patterns With Generalized Linear Regression Model. Front Genet 2019;10. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00120. Available at: Accessed May 29, 2023.
33. Byvatov E, Schneider G. Support vector machine applications in bioinformatics. Appl Bioinf 2003;2(2):67–77.
34. Chen X, Ishwaran H. Random Forests for Genomic Data Analysis. Genomics 2012;99(6):323–9.
35. Wu FX. Genetic weighted k-means algorithm for clustering large-scale gene expression data. BMC Bioinf 2008;9(6):S12.
36. Kobak. The art of using t-SNE for single-cell transcriptomics | Nature Communications. Available at: https://www.nature.com/articles/s41467-019-13056-x Accessed May 30, 2023.
37. Alipanahi B, Delong A, Weirauch MT, et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 2015;33(8):831–8.
38. Luo X, Tu X, Ding Y, et al. Expectation pooling: an effective and interpretable pooling method for

predicting DNA–protein binding. Bioinformatics 2020;36(5):1405–12.
39. Novakovsky G, Dexter N, Libbrecht MW, et al. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nat Rev Genet 2023;24(2):125–37.
40. Vaz JM, Balaji S. Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. Mol Divers 2021;25(3):1569–84.
41. Abdeltawab H, Khalifa F, Ghazal M, et al. A pyramidal deep learning pipeline for kidney whole-slide histology images classification. Sci Rep 2021;11(1):20189.
42. Schulte-Sasse R, Budach S, Hnisz D, et al. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. Nat Mach Intell 2021;3(6):513–26.
43. Zhang Z, Park CY, Theesfeld CL, et al. An automated framework for efficiently designing deep convolutional neural networks in genomics. Nat Mach Intell 2021;3(5):392–400.
44. Angermueller C, Lee HJ, Reik W, et al. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol 2017;18(1):67.
45. Pan X, Rijnbeek P, Yan J, et al. Prediction of RNA-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks. BMC Genom 2018;19(1):511.
46. Quang D, Xie X. DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. Nucleic Acids Res 2016;44(11):e107.
47. Quang D, Xie X. FactorNet: A deep learning framework for predicting cell type specific transcription factor binding from nucleotide-resolution sequential data. Methods San Diego Calif 2019;166:40–7.
48. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12(1):56–68.
49. Zhang XM, Liang L, Liu L, et al. Graph Neural Networks and Their Current Applications in Bioinformatics. Front Genet 2021;12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.690049. Available at: Accessed May 30, 2023.
50. Wang K, Li Z, You ZH, et al. Adversarial dense graph convolutional networks for single-cell classification. Bioinformatics 2023;39(2):btad043.
51. Tangherloni A, Ricciuti F, Besozzi D, et al. Analysis of single-cell RNA sequencing data based on autoencoders. BMC Bioinf 2021;22(1):309.
52. Doncevic D, Herrmann C. Biologically informed variational autoencoders allow predictive modeling of genetic and drug induced perturbations. biorxiv 2022. https://doi.org/10.1101/2022.09.20.508703.
53. Talwar D, Mongia A, Sengupta D, et al. AutoImpute: Autoencoder based imputation of single-cell RNA-seq data. Sci Rep 2018;8(1):16329.
54. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016;3(1):9.
55. Guo C, Wei B, Yu K. Deep Transfer Learning for Biology Cross-Domain Image Classification. J Control Sci Eng 2021;2021:e2518837.
56. Agarwal A, Chen L. DeepPHiC: predicting promoter-centered chromatin interactions using a novel deep learning approach. Bioinformatics 2023;39(1):btac801.
57. Rampášek L, Hidru D, Smirnov P, et al. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. Bioinformatics 2019;35(19):3743–51.
58. Grønbech CH, Vording MF, Timshel PN, et al. scVAE: variational auto-encoders for single-cell gene expression data. Berger B. Bioinformatics 2020;36(16):4415–22.
59. Lan L, You L, Zhang Z, et al. Generative Adversarial Networks and Its Applications in Biomedical Informatics. Front Public Health 2020;8. https://doi.org/10.3389/fpubh.2020.00164.
60. Wang X, Ghasedi Dizaji K, Huang H. Conditional generative adversarial network for gene expression inference. Bioinformatics 2018;34(17):i603–11.
61. Viñas R, Andrés-Terré H, Liò P, et al. Adversarial generation of gene expression data. Bioinformatics 2022;38(3):730–7.
62. Wang Z, Chen Z, Zhao H, et al. ISPRF: a machine learning model to predict the immune subtype of kidney cancer samples by four genes. Transl Androl Urol 2021;10(10):3773–86.
63. Jagga Z, Gupta D. Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms. BMC Proc 2014;8(Suppl 6):S2.
64. Marquardt A, Solimando AG, Kerscher A, et al. Subgroup-Independent Mapping of Renal Cell Carcinoma—Machine Learning Reveals Prognostic Mitochondrial Gene Signature Beyond Histopathologic Boundaries. Front Oncol 2021;11. Available at: https://www.frontiersin.org/articles/10.3389/fonc.2021.621278. Available at: Accessed May 27, 2023.
65. Terrematte P, Andrade DS, Justino J, et al. A Novel Machine Learning 13-Gene Signature: Improving Risk Analysis and Survival Prediction for Clear Cell Renal Cell Carcinoma Patients. Cancers 2022;14(9):2111.
66. Liu B, Xiao Y, Li H, et al. Identification and Verification of Biomarker in Clear Cell Renal Cell Carcinoma via Bioinformatics and Neural Network Model. BioMed Res Int 2020;2020:6954793.

67. Malouf GG, Su X, Zhang J, et al. DNA Methylation Signature Reveals Cell Ontogeny of Renal Cell Carcinomas. Clin Cancer Res Off J Am Assoc Cancer Res 2016;22(24):6236–46.
68. Pirmoradi S, Teshnehlab M, Zarghami N, et al. A self-organizing deep neuro-fuzzy system approach for classification of kidney cancer subtypes using miRNA genomics data. Comput Methods Programs Biomed 2021;206:106132. https://doi.org/10.1016/j.cmpb.2021.106132.
69. Ng KL, Taguchi YH. Identification of miRNA signatures for kidney renal clear cell carcinoma using the tensor-decomposition method. Sci Rep 2020;10(1):15149.
70. Iwamura H, Mizuno K, Akamatsu S, et al. Machine learning diagnosis by immunoglobulin N-glycan signatures for precision diagnosis of urological diseases. Cancer Sci 2022;113(7):2434–45.
71. Manzi M, Palazzo M, Knott ME, et al. Coupled Mass-Spectrometry-Based Lipidomics Machine Learning Approach for Early Detection of Clear Cell Renal Cell Carcinoma. J Proteome Res 2021;20(1):841–57.
72. Bhalla S, Chaudhary K, Kumar R, et al. Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. Sci Rep 2017;7:44997.
73. Chen L, Xiang Z, Chen X, et al. A seven-gene signature model predicts overall survival in kidney renal clear cell carcinoma. Hereditas 2020;157(1):38.
74. Singh NP, Bapi RS, Vinod PK. Machine learning models to predict the progression from early to late stages of papillary renal cell carcinoma. Comput Biol Med 2018;100:92–9.
75. Singh NP, Vinod PK. Integrative analysis of DNA methylation and gene expression in papillary renal cell carcinoma. Mol Genet Genomics 2020;295(3):807–24.
76. Ning Z, Pan W, Chen Y, et al. Integrative analysis of cross-modal features for the prognosis prediction of clear cell renal cell carcinoma. Bioinforma Oxf Engl 2020;36(9):2888–95.
77. Holdbrook DA, Singh M, Choudhury Y, et al. Automated Renal Cancer Grading Using Nuclear Pleomorphic Patterns. JCO Clin Cancer Inform 2018;2:1–12.
78. Ing N, Huang F, Conley A, et al. A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome. Sci Rep 2017;7(1):13190.
79. Azuaje F, Kim SY, Perez Hernandez D, et al. Connecting Histopathology Imaging and Proteomics in Kidney Cancer through Machine Learning. J Clin Med 2019;8(10):1535.
80. Cheng J, Han Z, Mehra R, et al. Computational analysis of pathological images enables a better diagnosis of TFE3 Xp11.2 translocation renal cell carcinoma. Nat Commun 2020;11(1):1778. https://doi.org/10.1038/s41467-020-15671-5.
81. Cheng J, Mo X, Wang X, et al. Identification of topological features in renal tumor microenvironment associated with patient survival. Bioinformatics 2018;34(6):1024–30. https://doi.org/10.1093/bioinformatics/btx723.