Computer Methods and Programs in Biomedicine 244 (2024) 107978

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Multimodal deep learning for personalized renal cell carcinoma prognosis: Integrating CT imaging and clinical data

Maryamalsadat Mahootiha [a,b,*], Hemin Ali Qadir [a], Jacob Bergsland [a], Ilangko Balasingham [a,c]

[a] The Intervention Centre, Oslo University Hospital, Oslo, 0372, Norway
[b] Faculty of Medicine, University of Oslo, Oslo, 0372, Norway
[c] Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Radiological imaging
Renal cell carcinoma
Deep learning
Survival analysis
Cancer prognosis
ISUP grading

ABSTRACT

Background and Objective: Renal cell carcinoma represents a significant global health challenge with a low survival rate. The aim of this research was to devise a comprehensive deep-learning model capable of predicting survival probabilities in patients with renal cell carcinoma by integrating CT imaging and clinical data and addressing the limitations observed in prior studies. The aim is to facilitate the identification of patients requiring urgent treatment.

Methods: The proposed framework comprises three modules: a 3D image feature extractor, clinical variable selection, and survival prediction. Based on the 3D CNN architecture, the feature extractor module predicts the ISUP grade of renal cell carcinoma tumors linked to mortality rates from CT images. Clinical variables are systematically selected using the Spearman score and random forest importance score as criteria. A deep learning-based network, trained with discrete LogisticHazard-based loss, performs the survival prediction. Nine distinct experiments are performed, with varying numbers of clinical variables determined by different thresholds of the Spearman and importance scores.

Results: Our findings demonstrate that the proposed strategy surpasses the current literature on renal cancer prognosis based on CT scans and clinical factors. The best-performing experiment yielded a concordance index of 0.84 and an area under the curve value of 0.8 on the test cohort, which suggests strong predictive power.

Conclusions: The multimodal deep-learning approach developed in this study shows promising results in estimating survival probabilities for renal cell carcinoma patients using CT imaging and clinical data. This may have potential implications in identifying patients who require urgent treatment, potentially improving patient outcomes. The code created for this project is available for the public on: GitHub

1. Introduction

1.1. Overview

Renal cell carcinoma (RCC) is a common adult malignancy, accounting for approximately 90% of all kidney tumors [40]. It originates in the tubules responsible for filtering blood and producing urine within the kidney [40]. If RCC is not identified and addressed promptly, it can spread to other parts of the body, such as the lungs and bones, posing significant life-threatening risks [47]. The increasing global incidence of RCC could be due to advancements in diagnostic techniques, more frequent use of medical imaging, and shifts in lifestyle factors [44,61]. Early treatment of RCC is essential for optimizing patient outcomes, enhancing survival rates, and improving the quality of life [61].

Survival analysis is a statistical method designed to explore the time interval leading up to a pivotal event, like death or the recurrence of a disease, and is frequently employed in oncology. This approach evaluates time-to-event data, estimating the likelihood of an event transpiring within a defined timeframe, all while considering censored data. Crucially, survival analysis can incorporate participants who didn't encounter the event of interest by the study's conclusion [26].

Survival analysis holds significant importance for RCC patients, guiding treatment choices and allowing clinicians to determine the best strategies—this includes decisions about therapy types, treatment intensity, and the consideration of palliative or supportive care measures [20]. Radiological data plays a crucial role in cancer patients' survival analysis and prognosis. It provides insights into tumor features, heterogeneity, treatment planning, and evaluation of therapeutic responses.

---

* Corresponding author at: The Intervention Centre, Oslo University Hospital, Oslo, 0372, Norway.
E-mail address: marymaho@uio.no (M. Mahootiha).

https://doi.org/10.1016/j.cmpb.2023.107978
Received 1 June 2023; Received in revised form 5 December 2023; Accepted 12 December 2023

Such data empowers clinicians to enhance patient outcomes and survival rates [25]. However, even experts can occasionally make incorrect predictions or misinterpret medical imagery, leading to potential misjudgments in prognosis and treatment. Astonishingly, an estimated 20 million radiology reports annually are believed to contain significant clinical errors [4]. Added to this is the possible scarcity of skilled radiologists in some regions or healthcare systems. Introducing artificial intelligence (AI) technologies may offer a solution to these challenges [28].

AI offers a promising avenue to enhance the precision and speed of medical image analysis. Convolutional neural networks (CNN) can discern patterns and features in images that might elude human observation [9]. These sophisticated algorithms can swiftly and accurately analyze vast volumes of data, thereby diminishing human error and boosting diagnostic accuracy [30]. Furthermore, the integration of AI into survival analysis has demonstrated potential, with capabilities to refine the accuracy of prognostic models and pave the way for tailored treatments [53].

In this study, we aim to develop a multimodal, AI-driven algorithm that predicts personalized survival probabilities by leveraging both CT images and clinical data. This approach addresses the challenges of potential inaccuracies in clinical interpretations and the limited availability of expert radiologists. Our method focuses on a multimodal survival analysis strategy, striving for greater precision in predicting survival outcomes. To achieve this, we categorize RCC tumors in CT scans based on the International Society of Urological Pathology (ISUP) grading system [46]. This grading evaluates the severity of cancer by analyzing the morphological attributes of tumor cells under a microscope and has a proven correlation with mortality rates [41]. Following tumor classification, we extract radiomic features, which serve as our survival model's primary inputs. This model also integrates vital clinical data specific to each patient. By synergizing radiomic features with clinical data, we aim to derive survival probabilities using a non-linear and non-proportional approach, ensuring a more comprehensive, realistic, and precise survival estimation.

1.2. Related work

In the realm of statistics, the Cox proportional hazards (CPH) model [11] stands as a benchmark for conducting survival analysis with censored observations. However, its linear nature constrains its ability to identify non-linear relationships between input variables and event risks, such as death. With the rise of AI and deep learning (DL), there's now a broader scope for modeling survival analysis that embraces these non-linear dynamics. Models like Cox-nnet [8] and DeepSurv [22] have emerged to counteract CPH's limitations and uncover new prognostic indicators. Nevertheless, even these DL-based models grapple with the inherent proportional hazards assumption intrinsic to CPH. This assumption, which suggests a patient's covariates' influence on death risk remains consistent over time, might not always echo real clinical scenarios. This can produce survival curves that fail to intersect, presenting a potential misalignment with true patient outcomes.

Your paragraph is detailed and clearly outlines the innovative statistical methods. Here's a revised version for enhanced clarity and flow: Advancements in statistical modeling have introduced groundbreaking solutions to circumvent the limitations of the CPH model in survival analysis. Two notable methodologies designed to address the inherent linearity and proportionality constraints of CPH are multivariate time-to-event logistic regression (MTLR) [13] and Nnet-survival [15].

MTLR expands upon logistic regression to accommodate time-to-event data, modeling the joint probability of various events. This technique can integrate time-dependent covariates and adeptly manage non-proportional hazards, marking its significance in survival analysis. Conversely, Nnet-survival calculates each specific time interval's discrete conditional hazard rate. Though this concept has historical roots [5], its recent application in modern deep learning techniques has culminated in the creation of Nnet-survival. This method paves the way for individualized, non-proportional hazard probability curves tailored to each patient.

Multimodal deep learning (DL) [33] is a framework that harnesses DL methodologies to process and learn from varied data modalities, encompassing tabular data, images, and audio. In medical domains, where diverse datasets like clinical details, radiological images, and medication histories are prevalent, multimodal DL offers a robust platform. This approach makes it feasible to unravel intricate relationships between the diverse input data and the predicted outputs.

Previous studies have employed various approaches to conduct survival analysis, focusing on using radiological images or integrating radiological images with clinical variables to enhance survival estimation. Mukherjee et al. [31] developed a shallow CNN in conjunction with Cox loss to predict the prognosis of lung cancer patients using computed tomography (CT) image data alone. Wang et al. [54] presented a CNN autoencoder-based survival model incorporating Cox loss for predicting recurrence in patients with high-grade serous ovarian cancer, relying solely on CT scans. Wu et al. [57] developed a regression-based survival model for non-small cell lung cancer patients, effectively integrating imaging and clinical data to enhance the accuracy of survival predictions by employing the mean squared error (MSE) loss function. Zhang et al. [58] introduced a risk prediction model for assessing overall survival in gastric cancer patients, incorporating both CT images and clinical variables as inputs and utilizing a specialized loss function. Zhong et al. [59] presented a CNN-based model using Cox survival loss to predict survival outcomes in patients diagnosed with stage T3N1M0 nasopharyngeal carcinoma using magnetic resonance (MR) imaging and clinical variables. Lastly, Chaddad et al. [6] explored the potential of radiomic features and clinical variables in predicting the survival group of lung cancer patients. The authors employed image analysis techniques, rather than DL methods, to extract radiomic features and utilized a random forest classifier.

In renal cell carcinoma research, numerous efforts have been made to develop survival predictive models using clinical data, radiological imaging, and histopathology. Pignot et al. [36] utilized clinical variables such as age, gender, TNM stage, tumor size, and histologic subtype in Kaplan-Meier and Cox regression analyses, aiming to ascertain the prognostic utility of subtyping for papillary RCC. Similarly, Arrontes et al. [3] evaluated comorbidity indices by inputting clinical data, including age, gender, and Fuhrman grading, into Kaplan-Meier and Cox models, investigating their influence on the prognosis of clear cell RCC. Meanwhile, Gao et al. [14] constructed a Cox regression model by incorporating CT texture features derived from radiomic feature extraction to predict survival rates in renal cell carcinoma cases. In another approach, Nazari et al. [32] developed a random forest classifier that uses quantitative radiomic features from CT scans to estimate the 5-year mortality risk for clear cell RCC patients. Furthering this line of research, Jiang et al. [21] trained a random forest model on CT radiomic data to categorize patients into SSIGN (stage, size, grade, and necrosis) risk groups specifically for clear cell RCC. Tabibu et al. [48] harnessed a deep Residual Neural Network to process histopathology images, extracting histopathological image features via deep learning. These features were subsequently integrated with a Cox model to conduct a comprehensive survival analysis for pan-renal cell carcinoma. Ning et al. [34]) utilized sophisticated machine learning methods in conjunction with Cox proportional hazards to devise a survival model that integrated gene data and histopathological images. Two studies attempted to utilize deep learning for prognosis based on different data modalities and the Cox proportional hazards model [35,43]. Specifically, Ning et al. [35] integrated CT scans, histopathological images, clinical data, and genomic data, while Schulz et al. [43] used CT/MR images, histopathological images, and genomic data.

1.3. Our contributions

Key contributions of our model include:

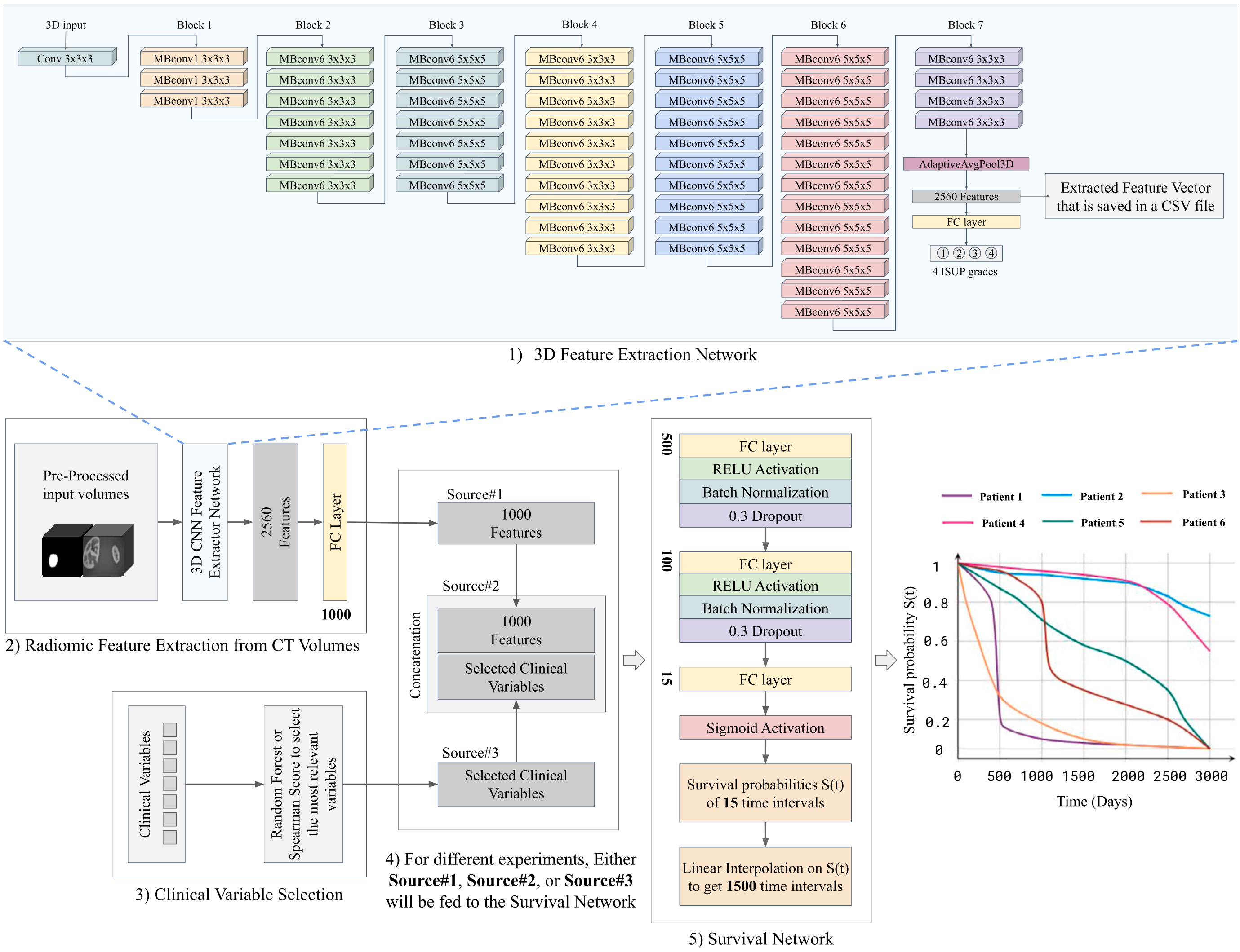

Fig. 1. Illustration of our comprehensive framework. (1) Feature Extraction from CT Volumes: Classifies CT images based on ISUP grades and utilizes a 1000-neuron fully connected layer to refine radiomic feature size from 2560. (2) Clinical Variables Selection: Merges CT image features and clinical variables using Spearman correlation and random forest importance scores. (3) Survival Network: Uses inputs from CT image features, clinical variables, and their combination to output survival probabilities for 15-time intervals, which are expanded to 1500 time points for patient-specific survival curve visualization.

1. Utilization of 3D inputs and convolutional layers to preserve detailed image features as inputs to the survival model.
2. Capability to perform non-proportional survival analyses for renal cell carcinoma, enhancing clinical applicability.
3. Leveraging deep learning survival analysis to validate the connection between ISUP grading and mortality rate.
4. Data combination selection mechanism to improve survival prediction accuracy.
5. Incorporation of violin diagrams for a nuanced evaluation of survival probability distributions.

2. Methods

As depicted in Fig. 1, our methodology for survival analysis integrates two primary data modalities: 1) CT volumes and 2) clinical variables. Drawing inspiration from the triumphant application of CNNs in image analysis and cancer prognosis, we incorporate a CNN-based architecture specifically tailored for CT image feature extraction pertinent to prognosis. Harnessing the capabilities of 3D CNNs, we pull out features from the three-dimensional tumor volume, a strategy influenced by Zhu et al. [60]. We then combine clinical information with the CT image features for survival analysis. Our approach consists of three modules: (1) CT image feature extraction, (2) selection of clinical variables, and (3) survival prediction. In our research, the feature extractor and survival networks undergo separate training processes rather than being trained simultaneously.

2.1. Radiomic feature extraction from CT volumes

We recommend classifying RCC tumors in CT images into ISUP grades (1, 2, 3, and 4) to extract radiomic features pertinent to prognosis. The CT volumes are processed using a 3D CNN feature extractor network to derive these features. Subsequently, we combine the clinical variables with the extracted radiomic features. Our choice of the ISUP grade for classification is based on its proven strong correlation with tumor recurrence, metastasis, and mortality [55]. Specifically, higher ISUP grades signal a poorer prognosis and increased mortality rate, while lower grades indicate a better prognosis and reduced mortality rate [10].

In our study, we have chosen EfficientNet [49] for the feature extractor network and classifier. Developed by Google researchers, this state-of-the-art CNN architecture is designed specifically for image classification. It employs the compound coefficient method, ensuring efficient model scaling. Among its variants, the largest model, EfficientNet

B7, stands out with superior performance. The layers of EfficientNet incorporate MBConv [42], an innovative convolutional block. This block captures intricate image features while utilizing fewer parameters and demanding less computational power compared to conventional convolutional blocks.

For our study, we modified the EfficientNet B7 architecture to cater to three-dimensional (3D) image data, such as CT volumes. This adaptation enables feature extraction across all dimensions of the tumor volume, ensuring that the spatial information inherent in 3D data is fully leveraged. Thus, our feature extraction network operates in a 3D domain. The input for this network consists of preprocessed image volumes combined with annotations from tumor segmentations. The purpose of this network is to classify RCC tumors into one of the four ISUP grades. We've conducted a detailed study separately on classifying RCC based on the ISUP grading system [29]. The architecture seamlessly blends convolutional layers, MBConv layers, an Adaptive Average Pooling layer, followed by a sequence of fully connected (FC) layers.

The Adaptive Average Pooling layer, positioned between the CNN and FC layers, is pivotal for feature extraction. It minimizes the parameters and computational demands of classification and retains vital image feature information [39]. From this layer, we derive feature vectors for each patient. After flattening the output of the Adaptive Average Pooling layer, the extracted image features are transformed into feature vectors. Initially, these vectors have a dimension of 2560. Our aim is to pare this down to 1000 for seamless integration with clinical data. We use an FC layer designed to take 2560 input features and produce 1000 output features to achieve this. These refined vectors are then saved in a CSV file, ready for input into the survival network. Once stored, we normalize the data, aligning it based on its mean and standard deviation.

2.2. Clinical variables selection

Our aim isn't to integrate all clinical variables with CT image features for survival prediction. Instead, we're interested in the potential of using a select subset of variables, specifically those most pertinent to prognosis, together with CT image features for better survival prediction outcomes. To achieve this, we assess different combinations of clinical variables. To pinpoint the most relevant clinical variables for predicting survival durations, we utilize two established techniques: (1) the Spearman correlation score [37] and (2) the importance score from a random forest regressor [56]. These methods aid us in recognizing the most valuable clinical variables to incorporate into our survival model, ensuring more precise patient survival predictions.

We chose random forest importance scores and Spearman rank correlation coefficients for feature selection for several reasons. First, they provide straightforward rankings of features based on their relevance, simplifying identifying the most important variables. Second, these methods quantify the relationship between each feature and survival time, elucidating how the variables associate with the outcome. Third, random forest and Spearman calculations are computationally efficient, particularly with high-dimensional data, enabling quick feature screening. Fourth, they use model-based and model-free evaluations to determine how important each feature is. This way, they can capture how important the feature is to the ensemble model and how it directly affects survival.

Spearman's rank correlation coefficient is a non-parametric statistical measure that assesses the strength and direction of the association between two variables. Unlike the parametric Pearson correlation coefficient, it is adept at identifying monotonic relationships and is less affected by non-linear relationships and non-normal distributions. We compute the Spearman correlation coefficients to gauge the relationship between clinical variables and survival times, resulting in a correlation matrix. This matrix showcases the pairwise correlation values between each clinical variable and survival times. The range of these coefficient values spans from -1 to 1: a value of -1 indicates a strong negative correlation, 1 signals a strong positive correlation, and 0 suggests no correlation.

Conversely, random forest regression can yield an importance score for each predictor variable. To determine these scores, we train a random forest model composed of 100 decision trees, aiming to predict survival times using clinical variables. These importance scores stem from the average reduction in the model's prediction error attributed to each feature, aggregated across all trees within the random forest. Following this, the clinical variables are ranked based on their respective importance scores to identify those most critical for survival time prediction. Variables with higher importance scores typically influence the model's predictive accuracy more.

2.3. Modeling survival estimation

This subsection delves into the nuances of modeling survival estimation, a cornerstone of our proposed method. The discussion is segmented into three distinct parts:

1. **Survival Network:** This segment outlines the architecture and design choices underpinning the survival network tasked with computing survival probabilities.
2. **Input to the Survival Network:** Here, we elaborate on the data and features serving as inputs to the network, encompassing clinical variables and radiomic features sourced from the 3D CNN feature extractor.
3. **Loss Function for Modeling Survival Estimation:** This portion highlights our choice of the loss function, which is pivotal in optimizing the survival network.

2.3.1. Survival network

The survival network, shown in Fig. 1, consists of three FC layers, comprising 500, 100, and 15 neurons, respectively. As our model is a discrete-time survival model, the final layer contains 15 neurons representing survival probabilities for 15 distinct time intervals. The network utilizes a rectified linear unit (ReLU) activation function in the intermediate layers and a sigmoid activation function in the last layer. In an effort to enhance the generalization capabilities of the model, a dropout rate of 0.3 is incorporated, accompanied by the implementation of batch normalization after the initial two FC layers. Subsequently, linear interpolation with 100 points is employed to transform the outputs into a set of 1500 values, enabling the generation of continuous survival curves for patients. We achieve the optimal architecture through a grid search of hyperparameters to find the best evaluation metrics for survival analysis.

2.3.2. Input to the survival network

The inputs to the survival network are derived from one of three sources: CT image features, clinical variables, or a combination of CT image features and clinical variables. In this study, we do a series of nine experiments, each using one of these three sources for survival prediction. In Section 3.4, a full explanation of these experiments will be given.

2.3.3. Loss function for modeling survival estimation

We adapt our survival model loss function based on discrete logistic hazards similar to the loss used in Nnet survival [15] to predict survival probabilities over M days (weeks, months, or years), in which M is the maximum follow-up period. It is essential to convert continuous survival times into discrete intervals to employ the discretized hazard function. A judicious selection of appropriate time intervals is undertaken to discretize the continuous survival times, with the preferred choice being equidistant intervals. Subsequently, each observed survival time is allocated to its respective time interval, effectively transforming the continuous data into a discrete format. We developed a loss function that used a vectorized form of likelihoods for censored and uncensored patients. The loss function is given by:

$$\mathrm{L} = -\sum_{x=1}^{p}\sum_{i=1}^{n}\left(\begin{array}{l}\ln\left(1+\mathrm{surv}_s(x)(i)\cdot\left(\mathrm{surv}_{\mathrm{pred}}\,(x)(i)-1\right)\right)\\ +\ln\left(1-\mathrm{surv}_f(x)(i)\cdot\mathrm{surv}_{\mathrm{pred}}\,(x)(i)\right)\end{array}\right),$$

where p denotes the number of patients in a batch, and n represents the number of discrete time intervals (15). $\mathrm{surv}_{\mathrm{pred}}(x)(i)$ signifies the predicted outcome of the survival model for patient x at time interval i, which can be either 0 for a patient who died during interval i or 1 for a patient who remained alive in interval i. Each patient's death or censoring time, t, is determined based on the ground truth survival time given in a dataset. The ground truth vectors $\mathrm{surv}_{\mathrm{s}}$ and $\mathrm{surv}_{\mathrm{f}}$ for the survival model are of length n for every patient. Vector $\mathrm{surv}_{\mathrm{s}}$ corresponds to the time intervals when the patient survived, while vector $\mathrm{surv}_{\mathrm{f}}$ denotes the specific time interval when the death occurred. For uncensored patients in the time interval i:

$$\mathrm{surv}_s(x)(i) = \begin{cases} 1, & \text{if } t_x \geq t_i \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{surv}_f(x)(i) = \begin{cases} 1, & \text{if } t_{i-1} \leq t_x < t_i \\ 0, & \text{otherwise} \end{cases}$$

For censored patients in the time interval i:

$$\mathrm{surv}_s(x)(i) = \begin{cases} 1, & \text{if } t_x \geq \frac{1}{2}\left(t_{i-1}+t_i\right) \\ 0, & \text{otherwise} \end{cases}$$

and

$$\mathrm{surv}_f(x)(i) = 0.$$

The dot product within the loss function assesses the similarities between the predicted vector and the ground truth vector. We trained the survival networks with the help of pycox v0.2.0.3 library.[1]

3. Experimental setup

In this section, we delineate the experimental setup utilized for our research. Our approach is segmented into four primary stages: data acquisition and preprocessing, training the 3D CNN feature extractor network, training the survival network, and executing the experiments. We commence by detailing the datasets incorporated, discussing their attributes, origins, and the preprocessing measures undertaken. Subsequently, we shed light on the training procedures for both the 3D CNN feature extractor and the survival networks. Lastly, we delve into a description of the various experiments executed.

3.1. Experimental dataset

The selection of appropriate datasets and their preparation plays a crucial role in evaluating our proposed method. This subsection provides an overview of the dataset used in our experiments and the steps taken to prepare the data for our study. We have divided this subsection into the KiTS21 dataset, dataset splitting, and clinical data preparation. First, we discuss the KiTS21 dataset, its characteristics, and its source. Next, we describe the dataset-splitting process, explaining the rationale behind the chosen method and the proportions used for training, validation, and testing. Finally, we detail the clinical data preparation, including any necessary preprocessing and data normalization procedures.

3.1.1. KiTS21 dataset

We used the KiTS21 [18] dataset to train and test our proposed framework. The dataset comprises 300 patients who underwent either partial or complete nephrectomy for suspected kidney cancer between 2010 and 2020 at the M Health Fairview or Cleveland Clinic medical facility and includes both clinical data and CT scans with manually annotated kidneys and tumors (ground-truth labels). The primary objective of collecting this dataset was to apply segmentation algorithms.

We selected this dataset for its comprehensive clinical information, precise annotations, and ample subject numbers. The dataset contains three files, including CT scan volumes (NIFTI format), annotation volumes (NIFTI format), and clinical data (JSON format). The annotation volumes consist of manual segmentations of the kidneys, tumor(s), and cyst(s). In this study, we used 41 clinical variables from this JSON file. All critical clinical information, such as pathology results, is included in this file [19]. Notably, this data was originally obtained from the Cancer Imaging Archive in DICOM format, while the clinical data was provided in a single CSV file.

3.1.2. Dataset splitting

To establish a classifier network conducive to radiomic feature extraction for survival prediction, we refined the original dataset by omitting 56 patients absent of ISUP grade values. The resulting dataset encapsulated 244 patients, 32 having experienced death events and 212 with censored time. The maximum observation time was 3000 days (which refers to the M variable in Section 2.3.3), and the median observation time was 644 days. We performed three-fold cross-validation for the ISUP grading classification to create three subsets for training, validation, and testing. The division of the dataset into three folds was based on the number of deceased and censored patients to ensure that each subset contained the same proportion of deceased individuals. Each fold included 57% of the total dataset for training, 10% for validation, and 33% for testing. The training subset had 10% of patients who died, the validation subset had 33%, and the test subset had 13%. After dividing the dataset into three folds, we increased the samples in each train and validation subset by doing multiple augmentations (discussed in 3.2.1).

The optimal fold for the classification model was identified based on the F1-score, as detailed in Section 3.2.3. This chosen fold was then applied for training, validation, and testing of the survival network, except that augmented samples were not used. It's worth noting that while separate networks were employed for ISUP grade classification and survival analysis, they were trained using the same cohort of subjects across training, validation, and test datasets. This methodology was specifically chosen to avoid the potential pitfalls of using the classification network's training data as the validation or test dataset for the survival analysis network. Such overlap could lead to an inflated perception of the survival analysis network's efficacy due to increased accuracy in identifying ISUP grades from the training dataset.

3.1.3. Clinical data preparation

The clinical data employed in training the survival network encompassed 38 variables, categorized into continuous numerical or categorical types. Categorical variables, like gender, were converted into discrete numerical values for compatibility with the survival model. Meanwhile, continuous numerical variables, such as pathologic size, underwent normalization using the mean and standard deviation to ensure optimal processing by the survival model.

3.2. Training the 3D CNN feature extractor

In this subsection, we elaborate on training the 3D CNN feature extractor, a critical component in our proposed method. This subsection is divided into three parts: 1) preprocessing of CT image volumes, which is a necessary step before training the 3D CNN feature extractor to guarantee consistent input data and enhance the network's performance; 2) training details of the classifier, encompassing aspects such as the chosen loss function, number of epochs, optimizer, and learning rate; 3)

---

[1] https://github.com/havakv/pycox.

best fold selection for radiomic feature extraction, a crucial step following the training of the 3D CNN feature extractor, which involves selecting the optimal fold to ensure the highest quality features for the subsequent survival network.

3.2.1. Preprocessing of CT image volumes

Before commencing the preprocessing phase for CT volumes, image augmentations were implemented as a strategy to address the inherent imbalance in the dataset, as well as the paucity of training samples. Positional augmentation was applied using affine transformations with shear parameters (0, 0.5, 0). Random horizontal and vertical flipping was also applied. In addition, rotations of up to 90 degrees were performed. Random Gibbs noise was added with probability one and alpha = (0.6, 0.8) for noise augmentation. Random SpaceSpike noise was also added with probability one and intensity_range = (10,13). Finally, Gaussian noise was added with a mean of 0 and a standard deviation of 1.

Following the augmentation, the count of CT scan samples rose from an initial 244 to 2,000; each ISUP grade class contains 500 samples. Before the ISUP grade classification, image preprocessing is applied to improve the quality of the input images and their radiomic features for better interpretation of the input [38,1]. As recommended in the MIT challenge,[2] all volumes were resized to 128 × 128 × 128. We also resampled the volumes based on one millimeter isotropic voxel size, which has been recommended as a standard voxel size by previous studies in medical imaging [2,52]. Additionally, all volumes were reoriented to the RAS (Right, Anterior, and Superior) orientation, the most commonly used orientation in medical images [2,52,27]. We utilized intensity normalization based on the Z-score in medical imaging [38,50]. For kidney image and tumor segmentation, identical image preprocessing steps were employed, except that intensity normalization was not applied for tumor segmentation.

We incorporated a concatenation step within our image preprocessing pipeline to boost the 3D EfficientNet-B7 model's efficacy in detecting kidney tumors. This step involved merging extracted kidney images with their associated manual tumor segmentations, aiming to highlight the tumors' surface patterns [1]. Such image concatenation supplements the input volume with crucial tumor position and dimensions data. Training the model solely on kidney images, without factoring in tumor location, might cause it to recognize irrelevant features, compromising its performance on new, unseen data. Consequently, our concatenation approach is pivotal in enhancing the model's generalizability and precision.

3.2.2. Training details

To validate the robustness of the radiomic feature extractor network, we conducted three-fold cross-validation with three distinct train, validation, and test subsets while maintaining the same hyperparameters for each training iteration. For training the 3D CNN feature extractor, we used the ADAM optimizer [23] with a fixed learning rate of 1×10^{-4}, and 50 epochs were run to optimize the network parameters. In addition, we employed the Cross-Entropy loss given by:

$$L = -\sum_{i=1}^{n} t_i \times log(p_i), \tag{1}$$

t_i is the true ISUP class, and p_i is the softmax probability for the ith class, and n is the number of ISUP classes (4 in this study). The 3D feature extractor was trained using PyTorch v1.11.0 on a workstation equipped with an Nvidia GeForce RTX 3090 GPU, an AMD Ryzen 7 5800X 8-Core Processor, and 32 GB of RAM.

3.2.3. Best fold selection for radiomic feature extraction

We used precision, recall, and F-score to evaluate our feature extractor network, as these fundamental metrics are indispensable for assessing classification model performance.

Precision, also known as the positive predictive value, quantifies the fraction of true positives out of the total instances predicted as positive by the model. Mathematically, precision can be defined as:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}, \tag{2}$$

where TP denotes true positives and FP denotes false positives.

Recall, alternatively referred to as sensitivity or true positive rate, measures the fraction of true positive instances among the total number of actual positive instances within the dataset. Recall can be mathematically represented as:

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{3}$$

where FN denotes false negatives.

The F-score, specifically the F1-score, constitutes the harmonic mean of precision and recall, delivering a single metric that balances both measures. The F1-score is particularly advantageous in situations with uneven class distributions, as it accounts for the trade-off between precision and recall. The F1-score can be calculated using the following equation:

$$\text{F1-score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{4}$$

We calculated the average of four Precision, Recall, and F-scores we gained for each ISUP class. We repeated this process three times for each of our three folds, giving us three average Precision, Recall, and F-scores. The second fold, with an average F-score of 0.84, was the best and selected as our final radiomic feature extractor that can be used as the input for the survival network.

3.3. Training the survival network

In the present study, we used a total of 500 epochs for training the survival network. Early stopping was implemented with a patience level of 10 to prevent overfitting and enhance generalization. The model was optimized utilizing the Adam optimizer, accompanied by a learning rate 0.01. The optimal learning rate selection was determined using the Smith [45] method.

3.4. Experiments

In our study, to demonstrate the performance improvement of our proposed survival analysis framework, we conduct nine distinct experiments with different combinations of inputs. The first experiment involves solely CT image features, the second only involves clinical variables, and the third combines CT image features and clinical variables. The remaining six experiments are created by applying three distinct thresholds for each the Spearman correlation and the random forest regression importance score. The clinical variables are selected based on the thresholds in the last six experiments and then fed to the survival network. These experiments are then compared to evaluate their effectiveness in predicting survival outcomes. Further details on the results of these experiments will be presented in Section 4.2.

4. Results

In this section, we delineate the performance assessment of our survival model, the resultant experimental outcomes, and a comparison with prior related studies. This section is segmented into three components:

---

[2] http://6.869.csail.mit.edu/fa17/miniplaces.html.

1. **Metrics for Survival Model Performance Evaluation:** This subsection elucidates the evaluation metrics harnessed to gauge the prowess of our proposed survival model.
2. **Experimental Results from Nine Distinct Experiments:** We present and analyze the results from nine experiments conducted to evaluate our proposed methodology.
3. **Plotting Violin Diagram for Survival Distribution:** This subsection accentuates the visualization of survival distribution data via violin diagrams, offering an encompassing perspective on our findings.

4.1. Metrics for performance evaluation

To assess the performance of our survival model, we used three key metrics: the time-dependent concordance index (C^{td}), the integrated Brier score (IBS), and the cumulative dynamic area under the curve (AUC). C^{td} extends Harrell's concordance index [17], a widely utilized measure for evaluating the discriminative power of survival models. The time-dependent C-index is calculated as follows:

$$C^{td}(t) = \frac{concordant\ pairs\ at\ time\ t}{comparable\ pairs\ at\ time\ t} \tag{5}$$

A concordant pair is when the patient with a greater predicted survival probability at time t actually outlives the other patient. A comparable pair, on the other hand, is when one patient's survival time is known and can be compared at time t. The time-dependent C-index is specifically designed to address situations where a model's predictive accuracy may vary. The C^{td} range between 0 and 1, where values approaching 1 signify superior predictive accuracy, while those nearing 0.5 indicate the model possesses no greater discriminative power than random chance. It has been established that the concordance index is excessively optimistic, particularly with an increasing number of censored patients in the dataset [51].

The cumulative dynamic AUC [24] extends the conventional AUC metric, a prominent measure for assessing binary classification models. This extension specifically addresses censored data and time-varying predictions in survival analysis. The cumulative dynamic AUC at time t is calculated as:

$$AUC(t) = \int_0^t ROC(u)du, \tag{6}$$

where ROC(u) is the time-dependent receiver operating characteristic curve at time u; within this context, the cumulative dynamic AUC is computed for a designated time point t, quantifying the model's discriminatory capacity to distinguish subjects experiencing the event of interest by time t from those who do not. The cumulative dynamic AUC represents the area under the time-dependent Receiver Operating Characteristic (ROC) curve, which delineates the sensitivity (true positive rate) against 1-specificity (false positive rate) for different time points. Ranging from 0 to 1, the cumulative dynamic AUC reveals greater predictive accuracy as it approaches 1, while values nearing 0.5 indicate that the model's discriminatory power is no better than random chance.

IBS [16] is an evaluation metric in survival analysis, assessing the calibration and overall predictive accuracy of time-to-event models. Originating from binary classification, the Brier Score measures average prediction error, while the IBS extends this principle to time-to-event data by integrating the Brier Score over a specific time horizon. The IBS is calculated as follows:

$$IBS = \int_0^\tau BS(t)dt, \tag{7}$$

where $BS(t)$ is the Brier Score at time t:

$$BS(t) = \sum_{i:T_i \geq t} \left(f_i(t) - I(T_i = t)\right)^2 / N(t), \tag{8}$$

in this context, $f_i(t)$ represents the forecasted probability of an event for the i-th individual at time t. $I(T_i = t)$ denotes the binary indicator for the event status, and $N(t)$ signifies the count of individuals at risk at time t.

Within the scope of survival analysis, the Brier Score computes the squared difference between observed binary event statuses and predicted event probabilities at a given time. The IBS incorporates the temporal dimension by aggregating prediction errors across a series of time points, considering data censoring. Lower IBS values signify superior calibration and overall predictive accuracy of the survival model, with 0 denoting a flawless prediction and 1 indicating the poorest possible prediction.

In addition to standard metrics, we use violin plots, a novel approach, to observe survival model output distributions. This is the first study proposing the application of violin plots to evaluate survival models. High evaluation metrics may be misleading, as predicted survival probabilities may not match ground truth times of death. Violin plots serve as a valuable tool in visualizing model performance by exhibiting the distribution of predicted probabilities at the time of mortality for deceased individuals and the distribution of predicted probabilities at the ultimate time point for censored subjects. For example, a distribution approximating zero for deceased patients signifies satisfactory model training, which consequently yields probability predictions close to zero.

4.2. Experimental results

One of our studies aims to investigate the impact of various combinations of clinical variables on predicting survival outcomes in patients with RCC. Specifically, we seek to identify the clinical features that contribute most significantly to accurately predicting patients' survival times. Initially, we conducted two independent analyses to evaluate the effectiveness of CT image features and clinical data individually with respect to their impact on the performance of our survival model. Subsequently, we explore the impact of merging CT image features with various combinations of selected clinical variables on the performance of the survival model.

To this end, we developed nine distinct experiments (Exp). Table 1 shows the difference between these nine experiments regarding their inputs and thresholds for choosing the combination of clinical variables. Table 1 also reports the C-index and AUC obtained on the test subset from each experiment. We used the same survival network architecture in the nine experiments for a fair comparison. From experiment 4 to experiment 9, we applied different thresholds for the Spearman correlation score (S_score) and random forest regression importance score (I_score).

We adjusted three different thresholds for Spearman's correlation coefficient. As the threshold values decreased, we incorporated more clinical variables with weaker correlations to the patient survival time into the survival model. In contrast, we utilized three different thresholds for the importance score of the decision tree regressor. By lowering these threshold values, we gradually incorporated less important clinical variables in predicting survival times into the survival model.

According to Table 1, the best evaluation metrics were obtained in experiment 8, in which the C-index, AUC, and IBS are 0.84, 0.8, and 0.18, respectively. The inputs to experiment 8 are the following: CT images features, Localized Solid Tumor, Age at Nephrectomy, Congestive Heart Failure, Body Mass Index, Uncomplicated Diabetes Mellitus, Pathologic Size, Myocardial Infarction, Radiographic Size, Metastatic Solid Tumor, Hospitalization, Mild Liver Disease, Smoking History, Surgery Type, Gender, Tumor Histologic Subtype, Pathology T Stage, and Surgical Approach. Please refer to Appendix A for additional clinical variables employed in the other experiments.

To evaluate the effectiveness of the survival model, ten unique individuals from the test cohort were selected, of which five had deceased from RCC, and five had censoring time to event. Subsequently, the

Table 1
Differences of Experiments used for RCC survival analysis.

Exp	Inputs	Theresholds	C-index	AUC	IBS
Exp1	CT images Features		0.72	0.73	0.19
Exp2	38 clinical variables		0.72	0.74	0.17
Exp3	CT images Features 38 clinical variables		0.82	0.74	0.19
Exp4	CT images Features 4 clinical variables	$\lvert S\_score \rvert \geqslant 0.1$	0.79	0.76	0.22
Exp5	CT images Features 13 clinical variables	$\lvert S\_score \rvert \geqslant 0.05$	0.83	0.75	0.21
Exp6	CT images Features 30 clinical variables	$\lvert S\_score \rvert \geqslant 0.01$	0.81	0.77	0.17
Exp7	CT images Features 4 clinical variables	I_score ⩾ 0.1	0.77	0.74	0.21
Exp8	**CT images Features** **17 clinical variables**	**I_score⩾0.01**	**0.84**	**0.8**	**0.18**
Exp9	CT images Features 29 clinical variables	I_score ⩾ 0.001	0.84	0.76	0.19

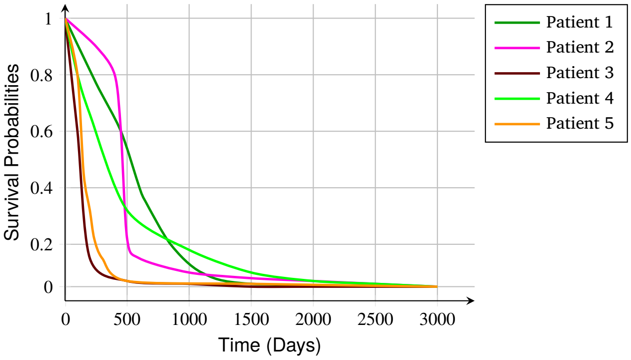

Fig. 2. Survival Probabilities for five patients in the test cohort who died.

survival curves for these patients were plotted, utilizing the survival probabilities derived from experiment 8. Fig. 2 illustrates five distinct survival curves generated by our survival model, corresponding to five different patients from the test cohort with events equal to one (deceased). Based on the ground truth survival time, patient 1 died after 645 days, patient two after 688 days, patient three after 102 days, patient four after 2,000 days, and patient five after 39 days.

At the time of their respective deaths, the model predicted survival probabilities of 0.42, 0.15, 0.3, 0.05, and 0.5 for patients 1 through 5. These values indicate varying degrees of accuracy in predicting the survival probabilities at the time of death, with patient 4 exhibiting the lowest probability and patient 5 the highest. At 500 days, the model's survival probability predictions for patients 1 to 5 were 0.57, 0.2, 0.06, 0.3, and 0.05, respectively. At 1000 days, these probabilities decreased to 0.1, 0.07, 0, 0.18, and 0 for the same patients. At 1500 days, all survival probability predictions reached 0, except for patient 4, whose probability reached 0 at 2000 days. The above findings suggest that the model demonstrates varying performance in predicting survival probabilities for the five patients at different time points. Some predictions align closely with the ground truth survival times, while others exhibit a bit of discrepancy.

Fig. 3 illustrates five distinct survival curves generated by our survival model for five different patients from the test cohort with events equal to zero (censored) and censoring times greater than 2000 days. Based on the ground truth survival time, their censoring times are 2473 days for patient 6, 2045 days for patient 7, 2900 days for patient 8, 2600 days for patient 9, and 2298 days for patient 10.

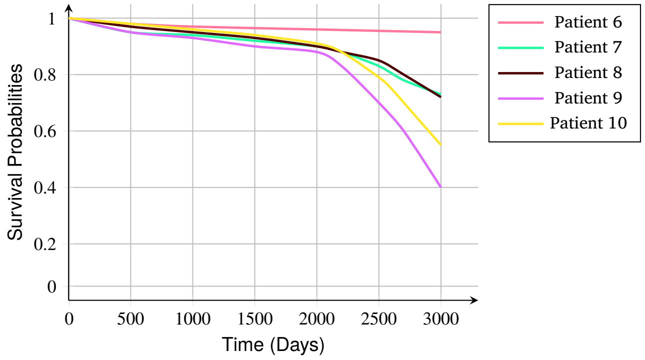

Fig. 3. Survival Probabilities for five patients in the test cohort who had censored events.

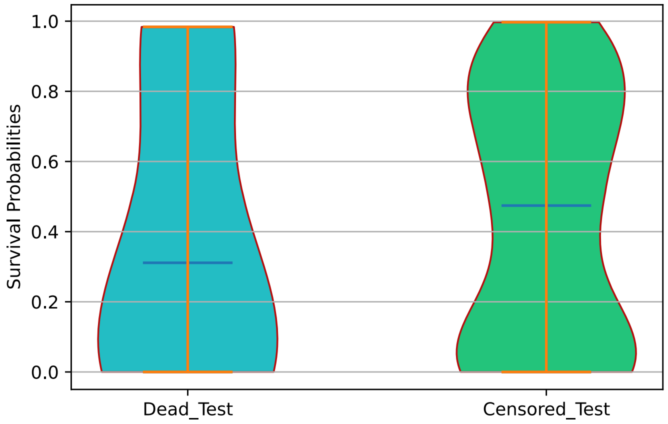

Fig. 4. Violin plots for censored & deceased events in train & test sets.

For patient 6, the model indicates a high probability of survival (0.95) at the censoring time of 2473, while patient 7 has a slightly lower survival probability of 0.9 at the censoring time of 2045. Patients 8, 9, and 10 exhibit survival probabilities of 0.75, 0.68, and 0.87 at their censoring times of 2900, 2600, and 2298, respectively. These predictions suggest that patient 6 has the highest likelihood of survival at their censoring time, followed by patients 7 and 10. Conversely, patients 8 and 9 possess relatively lower survival probabilities, with patient 9 exhibiting the lowest probability of survival among the five patients at their respective censoring times.

4.3. *Violin diagram for survival distribution*

Fig. 4 presents the violin plot for censored and uncensored subjects in the testing subset, showcasing the survival probability on the vertical axis for Exp8, which emerged as the optimal experimental outcome. As mentioned in Section 4.1, we can comprehend the distribution of survival probabilities predicted by our survival model with violin plots.

Censored_Test relates to the patients who did not experience the event in the test subset. Regarding the Censored_Test, we are uncertain about the outcomes at the final time (whether death occurred or not). Based on the median, it can be inferred that for half of the subjects, a survival probability lower than 0.45 would be predicted, with a higher concentration around 0.1. Conversely, a survival probability greater than 0.45 for the remaining half would be anticipated, with a greater distribution around 0.8. Given the symmetrical distribution around the median for Censored_Test, the model predicts that half of the censored patients would exhibit high survival probability at the last observation time. In contrast, the other half would demonstrate low survival probability. Dead_Test refers to patients who died within the test subset. This group's ideal output survival probabilities distribution is at zero. The median survival probability predicted by our survival model is around 0.3. Our survival model accurately predicted near-zero survival probabilities for half of the patients whose predicted probabilities were below the median. The other half of the patients with predicted probabilities higher than the median had distributions mostly near the median. Those nearer to the median had accurate survival predictions but with a small time shift. Those close to 1 are those patients whose survival probabilities were not accurately calculated. Upon analyzing the violin plots of the test subset for both censored and deceased patients, it can be concluded that our proposed multimodal survival model yields satisfactory outcomes that mostly align closely with the actual follow-up times of patients.

5. Discussion

In this section, we delve into the outcomes of our experiments, drawing comparisons with extant literature, examining the implications, our study limitations, and future research.

5.1. Validation of the study's hypothesis

The hypotheses underlying our study were twofold. Firstly, we aimed to investigate whether the selective provision of the most relevant clinical variables to the model would enhance the performance evaluation of survival analysis, as opposed to indiscriminately supplying all clinical variables. As evidenced by Table 1 in Section 4.2, our findings revealed that the most favorable results were obtained in Exp8, wherein clinical variables were judiciously chosen. In contrast, Exp 3, which included all clinical variables, yielded a lower C-index (by 0.02) and a reduced AUC (by 0.06). Our second hypothesis posited that multimodal survival analysis would yield superior results to single-modality approaches. In support of this hypothesis, Table 1 in Section 4.2 demonstrates that using single-modality data, such as solely clinical data or CT image features, led to lower performance metrics. In contrast, Exp 3 through 9, which incorporated clinical data and CT image features, significantly improved performance outcomes.

5.2. Survival curve analysis

To demonstrate that the integration of clinical data and CT image features results in superior performance compared to using CT image features or clinical data alone; we selected a single patient from the test cohort whose survival curve was incorrectly plotted in Exp 1 and Exp 2, in which both used a single data modality. This patient had an ISUP grade of 4 and a survival duration of 2,000 days. Subsequently, we generated survival curves for this patient from our nine defined experiments as illustrated in Fig. 5. The estimated survival probabilities for the selected patient at the time of death (2,000 days) were approximately 0.77 and 0.82 for Exp 1 and Exp 2, respectively. In contrast, the survival probabilities at the time of death for Exp3 through 9 were as follows: 0.18 for Exp3, 0.6 for Exp4, 0.61 for Exp5, 0.19 for Exp6, 0.55 for Exp7, 0.05 for Exp8, and 0 for Exp9. This result demonstrates that multimodal data can yield superior results to single-modality experiments.

5.3. Comparison with other studies

In the context of our study, we sought to draw comparisons with other studies that employed radiological images and clinical variables as inputs for their deep learning-based survival models to consider whether we could get acceptable performance metrics. A summary of these methodologies can be found in Section 1.2. Table 2 compares our approach with previous studies that used radiology imaging and clinical data as input and deep learning for survival analysis, focusing on the C-index and AUC metrics.

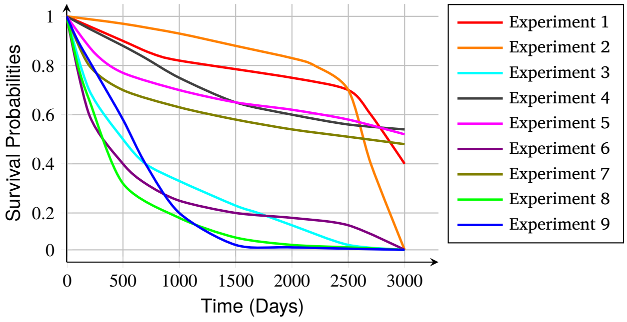

Fig. 5. Survival Probabilities from 9 different experiments for one patient.

We aimed to demonstrate that our results are within a similar range to prior work and the current state-of-the-art for multimodal prognosis modeling. Our methodology using 17 clinical variables achieved our study's highest C-index and AUC values. However, it is important to note that the prognostic factors differ biologically across cancer types and organs. Directly comparing the predictive performance of models across distinct diseases may not be appropriate, as the underlying disease biology and meaningful prognostic variables can vary. While our model showed strong predictive performance in renal cell carcinoma, the key prognostic factors and optimal modeling approaches are likely different for cancers in other organs.

As indicated in the second row of the Table 2 for Exp4, our approach's effectiveness remains evident even when only four clinical variables are employed. Additionally, it is worth noting that none of the aforementioned studies provided a methodology capable of generating non-proportional individualized survival curves for distinct patients. Furthermore, these studies relied on traditional methodologies susceptible to proportionality issues.

Table 2 presents a comparison of our method with other studies focusing solely on clinical data and radiological imaging for survival modeling, irrespective of cancer type. We also compared our approach to other methods in renal cell carcinoma prognosis that utilized various modalities as the inputs to their survival model as depicted in Table 3.

Ning et al. [35], Schulz et al. [43], and Ning et al. [34] methods are grounded in Cox loss, resulting in proportional survival curves that do not accurately mirror real-world clinical scenarios. Notably, Ning et al. [35] and Schulz et al. [43] studies that employed CT/MR imaging did not incorporate 3D images into their deep learning models. Ning et al. [35] extracted CT features using 2D scans and a patch-wise approach, while Schulz et al. [43] derived features from three separate images extracted from the entire 3D image, potentially overlooking crucial data linked to the interconnection of slices. However, we attempted to utilize 3D images as input to our 3D deep learning model for feature extraction. Our method intentionally avoids histopathological images, relying on CT scans and clinical data as non-invasive means of predicting survival probabilities.

While histopathological images offer comprehensive tumor information and were utilized in the referenced studies, it is noteworthy that our method yielded favorable outcomes even with a reduced number of modalities. It is crucial to clarify that the results of our model are derived from the analysis of the second fold. In contrast, the results for the other studies are aggregated from the mean values across multiple cross-validation folds. This distinction is important for accurately comparing and understanding the performance metrics presented.

5.4. Comparative analysis: MTLR vs. logistic hazard model

We intend to develop an MTLR model for Exp8 to understand if we can get better performance metrics with the MTLR model instead of the Logistic Hazard-based model. Table 4 shows the comparison between the Logistic hazard model and the MTLR model for Exp8.

Table 2
Comparison of this study results with previous related studies.

Studies	Organ	Cancer Type	Number of Clinical Variables	C-index	AUC
Our Method	Kidney	Renal Cell Carcinoma	17 (Exp8)	0.84	0.8
Our Method	Kidney	Renal Cell Carcinoma	4 (Exp4)	0.79	0.76
Chaddad et al. [6]	Lung	Non-Small Cell	2 (Age, TNM Stage)	-	0.76
Wu et al. [57]	Lung	Non-Small Cell	5 (Age, Histology, TNM Stage, Overall Stage, Gender)	0.65	-
Zhang et al. [58]	Stomach	Gastric	3 (Tumor Size, Tumor Localization, TNM Stage)	0.78	-
Zhong et al. [59]	Nasopharynx	Nasopharyngeal Carcinoma	3 (Age, LDH, Pre-EBV DNA)	0.78	-

Table 3
Comparison of this study results with related studies for renal cell carcinoma.

Studies	Input Modalities	C-index
Our Method	clinical data + CT	0.84
Ning et al. [35]	clinical data + histopathological images + genomic + CT	0.83
Schulz et al. [43]	histopathological images + genomic + CT/MR	0.77
Ning et al. [34]	histopathological images + genomic	0.71

Table 4
Comparison of survival analysis performance between MTLR and Logistic Hazard model for Exp8.

Model	C-index	AUC	IBS
Logistic Hazard	0.84	0.8	0.18
MTLR	0.75	0.73	0.18

Table 5
Comparison of survival analysis performance along different time interval numbers for Exp8.

Time Intervals	C-index	AUC	IBS
5	0.8	0.64	0.21
10	0.83	0.71	0.2
15	0.84	0.8	0.18
20	0.84	0.72	0.19
25	0.87	0.73	0.2
30	0.8	0.71	0.22

The MTLR model demonstrated inferior discriminative performance compared to the logistic hazard models, as evidenced by the lower C-index and AUC values. However, both models attained a comparable IBS. Given the decline in C-index and AUC with MTLR, these findings validate our initial methodology choice to utilize a logistic hazard regression framework for this discrete-time survival analysis task, as it demonstrated superior predictive accuracy over the alternative MTLR technique on this particular dataset.

5.5. *Effect of varying time intervals*

We aimed to modify the number of time intervals in our top-performing Experiment, Exp8, to observe variations in the model's performance. Table 5 shows the performance metrics for Exp8 with different time intervals as the model's output.

Fig. 6 illustrates our discrete model's IBS, C-index, and AUC values across different time intervals.

The IBS attained its minimum value of 0.18 using 15-time intervals, indicating optimal calibration at this discretization. Additionally, the AUC reached its peak value of 0.80 with 15 intervals, signifying maximum discriminative ability. While the C-index continued increasing up to 0.87 with 25 intervals, there was a notable gap between the C-index and AUC, which is 0.73 at this point, suggesting inflated performance. With 15 intervals, the C-index and AUC aligned closely at 0.84 and 0.80, respectively. Given the concordant optimization of both calibration as measured by IBS, and discrimination per the AUC, coupled with the tight agreement between the C-index and AUC, these results validate our selection of 15 discrete time intervals for the survival modeling methodology. The C-index improvement beyond 15 intervals is likely an artifact of overfitting, as evidenced by the divergence from the AUC. Thus, our empirical findings confirm that 15 intervals balance model performance and complexity for this dataset and modeling approach.

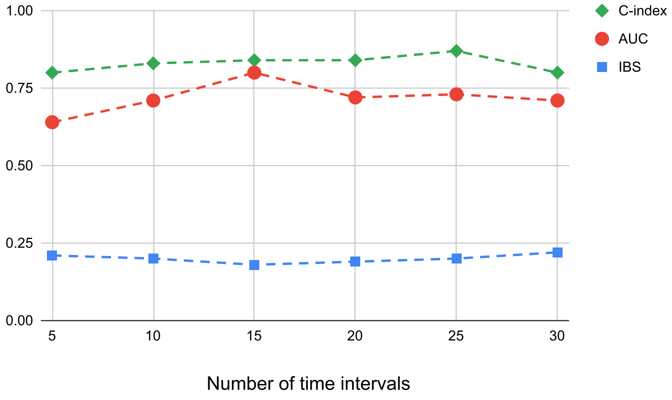

Fig. 6. Performance Metrics for different numbers of time intervals in Exp8.

5.6. *Impact of features from different ISUP classification folds*

Previously, we performed ISUP grade classification using three folds of data, where fold 2 showed the best performance. The radiomic features for survival analysis were extracted from the model trained on fold 2. To evaluate the effect of classification accuracy on prognosis prediction, we additionally extracted features from fold 1 and fold 3 models. This analysis was limited to experiment 1, which used only radiomic features to isolate the impact of the image-based ISUP grade classification on survival prediction. The features from fold 1 and 3 models were not as accurate due to poorer classification performance. Table 6 shows the survival analysis evaluation metrics obtained using radiomic features from the three classification folds. This allows us to compare prognosis prediction using features derived from models with varying ISUP grading accuracy. The results demonstrate the importance of accurate cancer grading for extracting prognostic radiomic signatures.

Table 6
Impact of features from different ISUP classification folds on survival analysis performance for Exp1.

Fold	F-score	C-index	AUC	IBS
1	0.6	0.65	0.63	0.24
2	0.84	0.72	0.73	0.19
3	0.72	0.67	0.70	0.15

We observed that higher ISUP grading F-scores correlated with improved survival analysis performance, as evidenced by higher C-index and AUC values. This supports our hypothesis that more accurate ISUP grading of tumor images enables the extraction of radiomic features that are more prognostic of patient outcomes. Fold 2 showed the best ISUP classification accuracy and survival analysis metrics when using radiomic features alone. Therefore, we determined that the radiomic features derived from the fold 2 classifier were optimal. These image-based features from fold 2 were selected to integrate with clinical variables in order to leverage the most accurate ISUP grading for survival prediction.

5.7. Study limitations

Despite the advantages of our methodology, our study presents several limitations. First, Experiment 8, which recorded the highest C-index and AUC, utilized 17 clinical variables during its training phase. For accurate survival predictions for a new patient, collecting all 17 clinical variables is imperative, as their entirety ensures precision in survival estimations. Second, accurate feature extraction demands complete abdomen images and segmentation annotations for the target organ and its related tumors. Third, to extend the findings of this study to other cancer types, it's crucial to identify a clinical variable analogous to the ISUP grade. This would facilitate tumor classification in context with survival estimations. Lastly, another limitation of our study is the high censoring in the dataset. This might impact the model's performance if we encounter new datasets with higher censoring levels, potentially hindering the achievement of similar performance results.

5.8. Future research

In future studies, we plan to merge RCC ISUP grade classification and survival prediction into a singular training framework, thereby eliminating the separate grading process for tumors. Moreover, we aim to delve into novel feature extraction techniques that obviate the need for organ and tumor annotations, further amplifying our proposed approach's effectiveness and broad applicability. We also plan to explore datasets with a broader range of data modalities, like histopathological images and genomic data, to evaluate how incorporating these additional modalities impacts the performance metrics of our model. Lastly, we can explore methodologies similar to Du et al. [12], Cheng et al. [7] studies to address the issue of high censoring in data inputs to our model.

6. Conclusion

This study presents a novel multimodal AI framework for predicting individualized survival probabilities in renal cell carcinoma patients. The proposed framework utilizes CT imaging and clinical data as inputs. We demonstrate that relevant features for survival estimation can be extracted from CT scans and combined with clinical variables to improve performance. Our framework generates personalized, nonlinear, non-proportional survival curves for individual patients, achieving higher accuracy than previously published methods. We show that a multimodal strategy enhances survival analysis accuracy compared to single-modality approaches. Moreover, we find that carefully selecting significant clinical factors as inputs further improves survival prediction performance. This work paves the path for enhanced clinical decision-making in renal cell carcinoma, enabling more precise, individualized therapy based on combined radiological and clinical data. Future research may build on these findings to create more sophisticated, reliable survival prediction models.

CRediT authorship contribution statement

Maryamalsadat Mahootiha: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hemin Ali Qadir:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Conceptualization. **Jacob Bergsland:** Writing – review & editing, Validation, Supervision. **Ilangko Balasingham:** Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

No conflict of interest exists.

Acknowledgement

The authors acknowledge the CIRCLE grant no. 287112 and the Health South-East Trust grant no. 2023069 for funding this study. We thank Håvard Kvamme, a previous Ph.D. student at the University of Oslo, for his invaluable guidance in effectively utilizing the pycox library he created.

Appendix A. Clinical variables

The following 38 clinical variables were included as covariates in the survival analysis in Experiment 2 and Experiment 3:

1. comorbidities, diabetes mellitus with end-organ damage
2. comorbidities, metastatic solid tumor
3. comorbidities, myocardial infarction
4. comorbidities, copd
5. comorbidities, uncomplicated diabetes mellitus
6. comorbidities, mild liver disease
7. comorbidities, hemiplegia from stroke
8. comorbidities, malignant lymphoma
9. comorbidities, congestive heart failure
10. comorbidities, localized solid tumor
11. comorbidities, chronic kidney disease
12. comorbidities, leukemia
13. comorbidities, dementia
14. comorbidities, connective tissue disease
15. comorbidities, cerebrovascular disease
16. comorbidities, peptic ulcer disease
17. comorbidities, peripheral vascular disease
18. comorbidities, moderate to severe liver disease
19. comorbidities, aids
20. intraoperative complications, injury to surrounding organ
21. chewing tobacco use
22. pathologic size
23. radiographic size
24. hospitalization
25. positive resection margins
26. smoking history
27. intraoperative complications, blood transfusion
28. pathology t stage
29. surgical procedure
30. age at nephrectomy
31. tumor histologic subtype
32. surgery type

33. gender
34. body mass index
35. malignant
36. surgical approach
37. cytoreductive
38. intraoperative complications, cardiac event

The following 4 clinical variables were included as covariates in the survival analysis in Experiment 4.

1. comorbidities, diabetes mellitus with end-organ damage
2. comorbidities, metastatic solid tumor
3. comorbidities, myocardial infarction
4. chewing tobacco use

The following 13 clinical variables were included as covariates in the survival analysis in Experiment 5.

1. comorbidities, diabetes mellitus with end-organ damage
2. comorbidities, metastatic solid tumor
3. comorbidities, myocardial infarction
4. comorbidities, copd
5. comorbidities, uncomplicated diabetes mellitus
6. comorbidities, mild liver disease
7. comorbidities, hemiplegia from stroke
8. comorbidities, malignant lymphoma
9. comorbidities, congestive heart failure
10. intraoperative complications, injury to surrounding organ
11. chewing tobacco use
12. pathologic size
13. radiographic size

The following 30 clinical variables were included as covariates in the survival analysis in Experiment 6.

1. comorbidities, diabetes mellitus with end-organ damage
2. comorbidities, metastatic solid tumor
3. comorbidities, myocardial infarction
4. comorbidities, congestive heart failure
5. comorbidities, uncomplicated diabetes mellitus
6. comorbidities, copd
7. comorbidities, mild liver disease
8. comorbidities, hemiplegia from stroke
9. comorbidities, malignant lymphoma
10. comorbidities, localized solid tumor
11. comorbidities, chronic kidney disease
12. comorbidities, leukemia
13. comorbidities, dementia
14. comorbidities, connective tissue disease
15. comorbidities, cerebrovascular disease
16. comorbidities, peptic ulcer disease
17. comorbidities, peripheral vascular disease
18. intraoperative complications, injury to surrounding organ
19. pathologic size
20. radiographic size
21. chewing tobacco use
22. hospitalization
23. positive resection margins
24. smoking history
25. intraoperative complications, blood transfusion
26. pathology t stage
27. surgical procedure
28. age at nephrectomy
29. tumor histologic subtype
30. surgery type

The following four clinical variables were included as covariates in the survival analysis in Experiment 7.

1. age at nephrectomy
2. body mass index
3. pathologic size
4. radiographic size

The following 29 clinical variables were included as covariates in the survival analysis in Experiment 9.

1. comorbidities, localized solid tumor
2. comorbidities, congestive heart failure
3. comorbidities, uncomplicated diabetes mellitus
4. comorbidities, myocardial infarction
5. comorbidities, metastatic solid tumor
6. comorbidities, mild liver disease
7. comorbidities, peripheral vascular disease
8. comorbidities, chronic kidney disease
9. comorbidities, diabetes mellitus with end-organ damage
10. comorbidities, peptic ulcer disease
11. comorbidities, copd
12. comorbidities, cerebrovascular disease
13. age at nephrectomy
14. body mass index
15. pathologic size
16. radiographic size
17. hospitalization
18. smoking history
19. surgery type
20. gender
21. tumor histologic subtype
22. pathology t stage
23. surgical approach
24. surgical procedure
25. cytoreductive
26. intraoperative complications, injury to surrounding organ
27. positive resection margins
28. intraoperative complications, blood transfusion
29. chewing tobacco use

References

[1] E. Akar, S. Kara, H. Akdemir, A. Kırış, Fractal analysis of mr images in patients with chiari malformation: the importance of pre-processing, Biomed. Signal Process. Control 31 (2017) 63–70, https://doi.org/10.1016/j.bspc.2016.07.005.

[2] M.Z. Alom, C. Yakopcic, M. Hasan, T.M. Taha, V.K. Asari, Recurrent residual U-net for medical image segmentation, J. Med. Imag. 6 (2019) 014006, https://doi.org/10.1117/1.JMI.6.1.014006.

[3] D.S. Arrontes, M.J.F. Aceñero, J.I.G. González, M.M. Muñoz, P.P. Andrés, Survival analysis of clear cell renal carcinoma according to the charlson comorbidity index, J. Urol. 179 (2008) 857–861, https://doi.org/10.1016/j.juro.2007.10.048.

[4] A.P. Brady, Error and discrepancy in radiology: inevitable or avoidable?, Insights Imaging 8 (2017) 171–182, https://doi.org/10.1007/s13244-016-0534-1.

[5] S. Brown, A. Branford, W. Moran, On the use of artificial neural networks for the analysis of survival data, IEEE Trans. Neural Netw. 8 (1997) 1071–1077, https://doi.org/10.1109/72.623209.

[6] A. Chaddad, C. Desrosiers, M. Toews, B. Abdulkarim, Predicting survival time of lung cancer patients using radiomic analysis, Oncotarget 8 (2017) 104393–104407, https://doi.org/10.18632/oncotarget.22251.

[7] X. Cheng, S. Wang, H. Wang, S.K. Ng, Deep survival forests for extremely high censored data, Appl. Intell. 53 (2023) 7041–7055, https://doi.org/10.1007/s10489-022-03846-0.

[8] T. Ching, X. Zhu, L.X. Garmire, Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data, PLoS Comput. Biol. 14 (2018) e1006076, https://doi.org/10.1371/journal.pcbi.1006076.

[9] F. Coppola, L. Faggioni, M. Gabelloni, F. De Vietro, V. Mendola, A. Cattabriga, M.A. Cocozza, G. Vara, A. Piccinino, S. Lo Monaco, L.V. Pastore, M. Mottola, S. Malavasi, A. Bevilacqua, E. Neri, R. Golfieri, Human, all too human? An all-around appraisal of the artificial intelligence revolution in medical imaging, Front. Psychol. 12 (2021) 710982, https://doi.org/10.3389/fpsyg.2021.710982.

[10] M. Costantini, M.L. Poeta, R.M. Pfeiffer, D. Hashim, C.L. Callahan, S. Sentinelli, L. Mendoza, M. Vicari, V. Pompeo, A.C. Pesatori, C.T. DellaValle, G. Simone, V.M. Fazio, M. Gallucci, M.T. Landi, Impact of histology and tumor grade on clinical outcomes beyond 5 years of follow-up in a large cohort of renal cell carcinomas, Clin. Genitourin. Cancer 19 (2021) e280–e285, https://doi.org/10.1016/j.clgc.2021.07.003.

[11] D.R. Cox, Regression models and life-tables, J. R. Stat. Soc. B 34 (1972) 187–202, https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

[12] D. Du, Q. Feng, W. Chen, Z. Ning, Y. Zhang, Mix-supervised multiset learning for cancer prognosis analysis with high-censoring survival data, Expert Syst. Appl. 239 (2024) 122430, https://doi.org/10.1016/j.eswa.2023.122430.

[13] S. Fotso, Deep neural networks for survival analysis based on a multi-task framework, https://doi.org/10.48550/arXiv.1801.05512, 2018.

[14] R. Gao, H. Qin, P. Lin, C. Ma, C. Li, R. Wen, J. Huang, D. Wan, D. Wen, Y. Liang, J. Huang, X. Li, X. Wang, G. Chen, Y. He, H. Yang, Development and validation of a radiomic nomogram for predicting the prognosis of kidney renal clear cell carcinoma, Front. Oncol. 11 (2021) 613668, https://doi.org/10.3389/fonc.2021.613668.

[15] M.F. Gensheimer, B. Narasimhan, A scalable discrete-time survival model for neural networks, PeerJ 7 (2019) e6257, https://doi.org/10.7717/peerj.6257.

[16] T.A. Gerds, M. Schumacher, Consistent estimation of the expected Brier score in general survival models with right-censored event times, Biom. J. 48 (2006) 1029–1040, https://doi.org/10.1002/bimj.200610301.

[17] F.E. Harrell, R.M. Califf, D.B. Pryor, K.L. Lee, R.A. Rosati, Evaluating the yield of medical tests, JAMA 247 (1982) 2543–2546, https://doi.org/10.1001/jama.1982.03320430047030.

[18] N. Heller, F. Isensee, K.H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, G. Yao, Y. Gao, Y. Zhang, Y. Wang, F. Hou, J. Yang, G. Xiong, J. Tian, C. Zhong, J. Ma, J. Rickman, J. Dean, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, H. Kaluzniak, S. Raza, J. Rosenberg, K. Moore, E. Walczak, Z. Rengel, Z. Edgerton, R. Vasdev, M. Peterson, S. McSweeney, S. Peterson, A. Kalapara, N. Sathianathen, N. Papanikolopoulos, C. Weight, The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge, Med. Image Anal. 67 (2021) 101821, https://doi.org/10.1016/j.media.2020.101821.

[19] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, C. Weight, The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes, https://doi.org/10.48550/arXiv.1904.00445, 2019.

[20] D. Hui, C.E. Paiva, E.G. Del Fabbro, C. Steer, J. Naberhuis, M. van de Wetering, P. ndez Ortega, T. Morita, S.Y. Suh, E. Bruera, M. Mori, Prognostication in advanced cancer: update and directions for future research, Support. Care Cancer 27 (2019) 1973–1984, https://doi.org/10.1007/s00520-019-04727-y.

[21] Y. Jiang, W. Li, C. Huang, C. Tian, Q. Chen, X. Zeng, Y. Cao, Y. Chen, Y. Yang, H. Liu, Y. Bo, C. Luo, Y. Li, T. Zhang, R. Wang, Preoperative ct radiomics predicting the ssign risk groups in patients with clear cell renal cell carcinoma: development and multicenter validation, Front. Oncol. 10 (2020), https://doi.org/10.3389/fonc.2020.00909.

[22] J.L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, Deepsurv: personalized treatment recommender system using a Cox proportional hazards deep neural network, BMC Med. Res. Methodol. 18 (2018), https://doi.org/10.1186/s12874-018-0482-1.

[23] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, https://doi.org/10.48550/arXiv.1412.6980, 2014.

[24] J. Lambert, S. Chevret, Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent roc curves, Stat. Methods Med. Res. 25 (2016) 2088–2102, https://doi.org/10.1177/0962280213515571.

[25] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J. Aerts, Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (2012) 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.

[26] E.T. Lee, J.W. Wang, Statistical Methods for Survival Data Analysis, John Wiley & Sons, Inc., 2003.

[27] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[28] X. Liu, L. Faes, A.U. Kale, S.K. Wagner, D.J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J.R. Ledsam, M.K. Schmid, K. Balaskas, E.J. Topol, L.M. Bachmann, P.A. Keane, A.K. Denniston, A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis, Lancet Digit. Health 1 (2019) e271–e297, https://doi.org/10.1016/S2589-7500(19)30123-2.

[29] M. Mahootiha, H.A. Qadir, J. Bergsland, I. Balasingham, Classification of kidney tumor grading on preoperative computed tomography scans, in: A. Tsanas, A. Triantafyllidis (Eds.), Pervasive Computing Technologies for Healthcare, Springer Nature, Switzerland, Cham, 2023, pp. 75–89.

[30] A. n Montero, U. Javaid, G. s, D. Nguyen, P. Desbordes, B. Macq, S. Willems, L. Vandewinckele, M. m, F. fman, S. Michiels, K. Souris, E. Sterpin, J.A. Lee, Artificial intelligence and machine learning for medical imaging: a technology review, Phys. Med. 83 (2021) 242–256, https://doi.org/10.1016/j.ejmp.2021.04.016.

[31] P. Mukherjee, M. Zhou, E. Lee, A. Schicht, Y. Balagurunathan, S. Napel, R. Gillies, S. Wong, A. Thieme, A. Leung, O. Gevaert, A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional CT-image data, Nat. Mach. Intell. 2 (2020) 274–282, https://doi.org/10.1038/s42256-020-0173-6.

[32] M. Nazari, I. Shiri, H. Zaidi, Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients, Comput. Biol. Med. 129 (2021) 104135, https://doi.org/10.1016/j.compbiomed.2020.104135.

[33] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: Proceedings of the 28th International Conference on International Conference on Machine Learning, Omnipress, 2011, pp. 689–696.

[34] Z. Ning, Z. Lin, Q. Xiao, D. Du, Q. Feng, W. Chen, Y. Zhang, Multi-constraint latent representation learning for prognosis analysis using multi-modal data, IEEE Trans. Neural Netw. Learn. Syst. 34 (2023) 3737–3750, https://doi.org/10.1109/TNNLS.2021.3112194.

[35] Z. Ning, W. Pan, Y. Chen, Q. Xiao, X. Zhang, J. Luo, J. Wang, Y. Zhang, Integrative analysis of cross-modal features for the prognosis prediction of clear cell renal cell carcinoma, Bioinformatics 36 (2020) 2888–2895, https://doi.org/10.1093/bioinformatics/btaa056.

[36] G. Pignot, C. Elie, S. Conquy, A. Vieillefond, T. Flam, M. Zerbib, B. Debré, D. Amsellem-Ouazana, Survival analysis of 130 patients with papillary renal cell carcinoma: prognostic utility of type 1 and type 2 subclassification, Urology 69 (2007) 230–235, https://doi.org/10.1016/j.urology.2006.09.052.

[37] W. Pirie, Spearman rank correlation coefficient, vol. 8, https://doi.org/10.1002/0470011815.b2a15150, 2006.

[38] F. Pérez-García, R. Sparks, S. Ourselin, Torchio: a python library for efficient loading, pre-processing, augmentation and patch-based sampling of medical images in deep learning, Comput. Methods Programs Biomed. 208 (2021) 106236, https://doi.org/10.1016/j.cmpb.2021.106236.

[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[40] A.M. Saad, M.M. Gad, M.J. Al-Husseini, I.A. Ruhban, M.B. Sonbol, T.H. Ho, Trends in renal-cell carcinoma incidence and mortality in the United States in the last 2 decades: a seer-based study, Clin. Genitourin. Cancer 17 (2019) 46–57.e5, https://doi.org/10.1016/j.clgc.2018.10.002.

[41] H. Samaratunga, T. Gianduzzo, B. Delahunt, The isup system of staging, grading and classification of renal cell neoplasia, J. Kidney Cancer VHL 1 (2014) 26, https://doi.org/10.15586/jkcvhl.2014.11.

[42] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[43] S. Schulz, A.C. Woerl, F. Jungmann, C. Glasner, P. Stenzel, S. Strobl, A. Fernandez, D.C. Wagner, A. Haferkamp, P. Mildenberger, W. Roth, S. Foersch, Multimodal deep learning for prognosis prediction in renal cancer, Front. Oncol. 11 (2021), https://doi.org/10.3389/fonc.2021.788740.

[44] R.L. Siegel, K.D. Miller, A. Jemal, Colorectal cancer statistics, CA Cancer J. Clin. 70 (2020) 7–30, https://doi.org/10.3322/caac.21601.

[45] L. Smith, Cyclical learning rates for training neural networks, https://doi.org/10.1109/WACV.2017.58, 2017, pp. 464–472.

[46] J.R. Srigley, B. Delahunt, J.N. Eble, L. Egevad, J.I. Epstein, D. Grignon, O. Hes, H. Moch, R. Montironi, S.K. Tickoo, M. Zhou, P. Argani, ISUP renal tumor panel, The international society of urological pathology (ISUP) Vancouver classification of renal neoplasia, Am. J. Surg. Pathol. 37 (2013) 1469–1489, https://doi.org/10.1097/PAS.0b013e318299f2d1.

[47] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[48] S. Tabibu, P.K. Vinod, C.V. Jawahar, Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning, Sci. Rep. 9 (2019) 10509, https://doi.org/10.1038/s41598-019-46718-3.

[49] M. Tan, Q.V. Le, Efficientnet: rethinking model scaling for convolutional neural networks, https://doi.org/10.48550/arXiv.1905.11946, 2019.

[50] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, J.C. Gee, N4ITK: improved N3 bias correction, IEEE Trans. Med. Imaging 29 (2010) 1310–1320, https://doi.org/10.1109/TMI.2010.2046908.

[51] H. Uno, T. Cai, M.J. Pencina, R.B. D'Agostino, L.J. Wei, On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data, Stat. Med. 30 (2011) 1105–1117, https://doi.org/10.1002/sim.4154.

[52] R. Vankdothu, M.A. Hameed, Brain tumor mri images identification and classification based on the recurrent convolutional neural network, Measurement: Sensors (2022) 100412, https://doi.org/10.1016/j.measen.2022.100412.

[53] P. Wang, Y. Li, C.K. Reddy, Machine learning for survival analysis: a survey, ACM Comput. Surv. 51 (2019), https://doi.org/10.1145/3214306.

[54] S. Wang, Z. Liu, Y. Rong, B. Zhou, Y. Bai, W. Wei, W. Wei, M. Wang, Y. Guo, J. Tian, Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer, Radiother. Oncol. 132 (2019) 171–177, https://doi.org/10.1016/j.radonc.2018.10.019.

[55] A.Y. Warren, D. Harrison, WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies, World J. Urol. 36 (2018) 1913–1926, https://doi.org/10.1007/s00345-018-2447-8.

[56] M. Wehenkel, A. Sutera, C. Bastin, P. Geurts, C. Phillips, Random forests based group importance scores and their statistical interpretation: application for Alzheimer's disease, Front. Neurosci. 12 (2018), https://doi.org/10.3389/fnins.2018.00411.

[57] Y. Wu, J. Ma, X. Huang, S.H. Ling, S. Weidong Su, Deepmmsa: a novel multimodal deep learning method for non-small cell lung cancer survival analysis, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 1468–1472.

[58] L. Zhang, D. Dong, W. Zhang, X. Hao, M. Fang, S. Wang, W. Li, Z. Liu, R. Wang, J. Zhou, J. Tian, A deep learning risk prediction model for overall survival in patients with gastric cancer: a multicenter study, Radiother. Oncol. 150 (2020) 73–80, https://doi.org/10.1016/j.radonc.2020.06.010.

[59] L.Z. Zhong, X.L. Fang, D. Dong, H. Peng, M.J. Fang, C.L. Huang, B.X. He, L. Lin, J. Ma, L.L. Tang, J. Tian, A deep learning mr-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage t3n1m0, Radiother. Oncol. 151 (2020) 1–9, https://doi.org/10.1016/j.radonc.2020.06.050.

[60] W. Zhu, C. Liu, W. Fan, X. Xie, Deeplung: deep 3d dual path nets for automated pulmonary nodule detection and classification, https://doi.org/10.1109/WACV.2018.00079, 2018, 673–681.

[61] A. Znaor, J. Lortet-Tieulent, M. Laversanne, A. Jemal, F. Bray, International variations and trends in renal cell carcinoma incidence and mortality, Eur. Urol. 67 (2015) 519–530, https://doi.org/10.1016/j.eururo.2014.10.002.