The American Journal of Pathology, Vol. 192, No. 6, June 2022

The American Journal of
PATHOLOGY
ajp.amjpathol.org

MACHINE LEARNING, COMPUTATIONAL PATHOLOGY, AND BIOPHYSICAL IMAGING

# Deep Learning of Rhabdomyosarcoma Pathology Images for Classification and Survival Outcome Prediction

Xinyi Zhang,* Shidan Wang,* Erin R. Rudzinski,† Saloni Agarwal,‡ Ruichen Rong,* Donald A. Barkauskas,§ Ovidiu Daescu,‡ Lauren Furman Cline,¶ Rajkumar Venkatramani,‖** Yang Xie,*†† Guanghua Xiao,*†† and Patrick Leavey¶

*From the Quantitative Biomedical Research Center,* Department of Population and Data Sciences, the Division of Pediatric Hematology and Oncology,¶ Department of Pediatrics, and the Department of Bioinformatics,†† University of Texas Southwestern Medical Center, Dallas, Texas; the Division of Anatomic Pathology,† Seattle Children's Hospital, Seattle, Washington; the Department of Computer Sciences,‡ University of Texas at Dallas, Richardson, Texas; the Department of Population and Public Health Sciences,§ Keck School of Medicine of the University of Southern California, Los Angeles, California; the Department of Pediatrics,‖ Baylor College of Medicine, Houston, Texas; and the Oncology Section,** Texas Children's Hospital, Houston, Texas*

Accepted for publication March 3, 2022.

Address correspondence to Guanghua Xiao, Ph.D., Quantitative Biomedical Research Center, Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390; or Patrick Leavey, M.D., Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75390. E-mail: guanghua.xiao@utsouthwestern.edu or patrick.leavey@utsouthwestern.edu.

Rhabdomyosarcoma (RMS), the most common malignant soft tissue tumor in children, has several histologic subtypes that influence treatment and predict patient outcomes. Assistance with histologic classification for pathologists as well as discovery of optimized predictive biomarkers is needed. A convolutional neural network for RMS histology subtype classification was developed using digitized pathology images from 80 patients collected at time of diagnosis. A subsequent embryonal rhabdomyosarcoma (eRMS) prognostic model was also developed in a cohort of 60 eRMS patients. The RMS classification model reached a performance of an area under the receiver operating curve of 0.94 for alveolar rhabdomyosarcoma and an area under the receiver operating curve of 0.92 for eRMS at slide level in the test data set ($n = 192$). The eRMS prognosis model separated the patients into predicted high- and low-risk groups with significantly different event-free survival outcome (likelihood ratio test; $P = 0.02$) in the test data set ($n = 136$). The predicted risk group is significantly associated with patient event-free survival outcome after adjusting for patient age and sex (predicted high- versus low-risk group hazard ratio, 4.64; 95% CI, 1.05–20.57; $P = 0.04$). This is the first comprehensive study to develop computational algorithms for subtype classification and prognosis prediction for RMS histopathology images. Such models can aid pathology evaluation and provide additional parameters for risk stratification. *(Am J Pathol 2022, 192: 917–925; https://doi.org/10.1016/j.ajpath.2022.03.011)*

Rhabdomyosarcoma (RMS) has an incidence rate of approximately 4.5 cases per 1 million children (about 350 cases per year in the United States)[1] and represents the most common soft tissue malignancy in children. With a propensity for myogenic differentiation, RMS is classified into four histologic subtypes: embryonal (eRMS), alveolar (aRMS), spindle cell sclerosing (scRMS), and pleomorphic (adult type).[2–4] The prognostic value of RMS histology has been widely validated despite inconsistent definitions over time.[2,5,6] Specifically, determination of aRMS histology, which carries a worse prognosis than eRMS, has been under debate between a requirement for >50% alveolar features or any evidence of alveolar features over the last 20 years.[7] Perhaps presence of *FOXO1* gene fusion may be a more significant predictor of outcome for aRMS.[8–10] Other histologic features, such as anaplasia, which previously might have been considered an adverse prognostic biomarker, may

Supported by the Children's Cancer Fund of Dallas (P.L.), the QuadW Foundation (P.L.), the NIH grants NCI National Clinical Trials Network (NCTN) Operations Center grant U10CA180886 (P.L.), NCTN SDC grant U10CA180899 (P.L.), Children's Oncology Group (COG) Biospecimen Bank grant U24CA196173 (P.L.), U01CA249245 (G.X.), R01GM140012 (G.X.), R01GM141519 (G.X.), R01DE030656 (G.X.), P30CA142543 (Y.X.), and R35GM136375 (Y.X.), and the Cancer Prevention and Research Institute of Texas grants RP190107 (G.X.) and RP180805 (Y.X.).

Disclosures: None declared.


https://doi.org/10.1016/j.ajpath.2022.03.011

actually be a surrogate marker for TP53 mutant tumors.[11] Such nuances to histologic subtype classification emphasize the role of expert and experienced pathologists and underline the subjective and time-intensive task.[11]

In recent years, the advancement in whole slide image scanning technologies has enabled the generation of high-resolution digital tissue images that retain environmental, cellular, and subcellular details. In parallel, the recent breakthroughs in artificial intelligence have provided opportunities to develop computer algorithms, especially deep learning algorithms, to aid tissue diagnosis.[12,13] Deep learning, of which convolutional neural networks (CNNs) are the most often implemented algorithms, is ideally suited for image classification tasks, such as the detection and segmentation of tumor regions,[13–16] to identification of lymph node metastases,[17–19] and to histologic grading, which in breast cancer is dependent on an evaluation of mitotic rate, nuclear morphology, and glandular differentiation.[20,21] CNNs generally have a series of convolution layers as hidden layers and thus constitute a deep network, which enables the extraction of representational image features to infer an output for prediction.[22,23] In this study, an RMS histology subtype classification model was developed on the basis of CNN to automatically classify image slides into three major subtypes of pediatric RMS, including eRMS, aRMS, and scRMS subtypes, as well as other tissues (mainly connective tissues) and background (mainly white space), as shown in Figure 1.

RMS risk stratification is based on several factors, including site of disease, surgical resectability, and tumor histology. Patients are classified as low, intermediate, or high risk based on risk of recurrence and death. However, despite advances in the biological understanding of this disease, the survival rate for children with high-risk and recurrent RMS remains unacceptably low at <20%. This failure in modern strategies highlights the need to optimize predictive biomarkers. Digitized histology images, analyzed using machine learning algorithms, provide a novel opportunity for optimization of biomarkers in a malignancy dependent on histologic subtype for stratification. However, one of the major challenges in developing a deep learning model for this purpose is the lack of adequate training samples in a rare disease. Transfer learning, which focuses on transferring the knowledge from a related domain to improve performance of another domain, is a promising solution for these machine learning problems and has been adopted for image analysis in brain tumors, colon cancer, and soft tissue sarcoma.[24–27] Typically, from a pretrained model for a specific task, the top one or several layer(s) for the classifier of the original task are retrained with the data set for a new task, whereas the remaining layers that play the role of feature extraction are kept untouched. Alternatively, the weights of a pretrained model can be used as initialization for the new task.[27] In this study, an eRMS prognostic model was developed by transfer learning from the RMS classification model, to separate eRMS patients into high- and low-risk groups.

This study shows one of the first pathology image-based models for histologic classification and prognosis prediction in pediatric RMS. Such models might be considered as an aid to pathology evaluation and might provide additional parameters for risk group stratification of patients with this rare cancer.[28]

## Materials and Methods

### Data Set

Children's Oncology Group approved a biobanking study, ARST18B4-Q, "Biomarker Development through Digital Histology Evaluated by Automated Learners in Combination with Genomic Characterization of Rhabdomyosarcoma," to study prospectively collected and digitized hematoxylin and eosin–stained pathology slides at the time of RMS biopsy and diagnosis. The pediatric RMS patients prospectively treated on a variety of Children's Oncology Group clinical trials had histology slides available from diagnostic biopsies; each histology slide representing a unique patient was digitized to a single whole slide image for analysis. In total, 340 whole slide images were reviewed and classified by a board-certified pathologist (E.R.R.) for this study. For simplification in this exploratory study, cases of RMS with anaplasia or sclerosing patterns, along with rare patterns like epithelioid or mixed RMS, were excluded. Thus, a total of 272 patient whole slide images with uniform appearance of a certain RMS subtype were included in this study, classified into aRMS ($n = 66$), eRMS ($n = 196$), and scRMS ($n = 10$). There was no pleomorphic RMS as this was a pediatric cancer data set. Clinical information and survival outcome for each patient were available, including sex, age, race, tumor stage and clinical group, tumor size, and *FOXO1* fusion status. The median follow-up time was 1830 days. Table 1 summarizes the patient characteristics of these 272 patients.

### Developing and Validating a Deep Learning Model for RMS Subtype Classification

Figure 2 summarizes the data sets, model development strategy (training, validation, and test), and results.

#### Training, Validation, and Testing Set Preparation

The RMS classification model was first developed to classify three major pediatric RMS histology subtypes at the image-patch level. From the cohort of 272 annotated slides, 80 slides (35 aRMS, 35 eRMS, and 10 scRMS) were chosen randomly to generate patch-level training, validation, and test data sets (Figure 2). Each of these 80 slides was annotated to identify five classes of different tissue types and regions representing aRMS, eRMS, scRMS, other tissue (mainly connective tissue, but also including necrosis and glands), and white background. Image patches with the

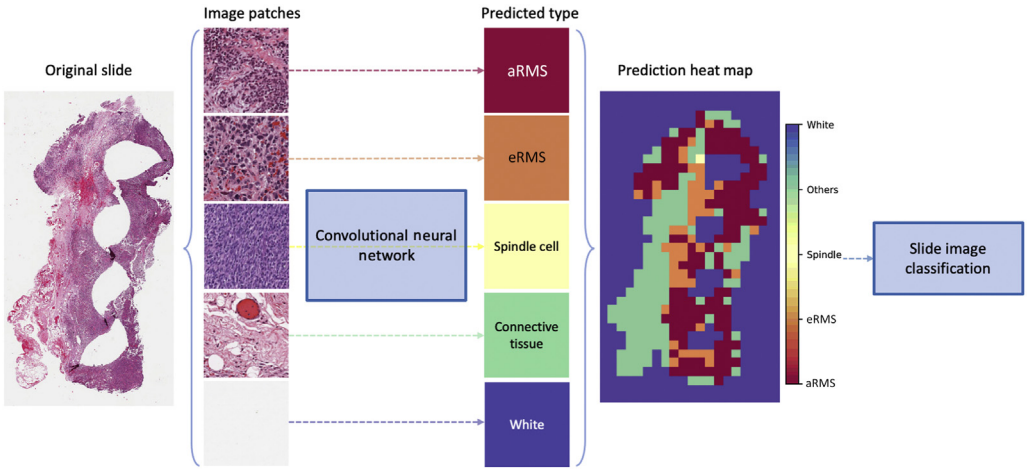



**Figure 1** Flowchart of rhabdomyosarcoma (RMS) classification model analysis process. aRMS, alveolar RMS; eRMS, embryonal RMS.

size of 500 × 500 pixels under ×20 magnification were extracted separately from these regions. Twenty-five aRMS slides were randomly assigned to generate 3107 training patches, five aRMS slides were randomly assigned to generate 876 validation patches, and five aRMS slides were assigned to generate the 513 test patches. The same distribution of slides was randomly assigned for the eRMS training, validation, and test data sets (25 eRMS slides were used to generate 3793 training patches, five eRMS slides were used to generate 570 validation patches, and five eRMS slides were used to generate 368 test patches). Given the smaller number of scRMS slides available, all 10 slides were used in the model development, where six scRMS slides were randomly assigned to generate 1246 image patches for training, two scRMS slides were used to generate 200 validation patches, and two scRMS slides were used to generate 203 test patches. Some examples of the training image patches are shown in Figure 3. The residual 128 slides were used as a slide-level test data set for whole slide classification.

### Patch-Level RMS Classification Model Training

InceptionV3 CNN architecture was adopted for the RMS classification model to classify image patches from five different tissue and region types, with input size 500 × 500 and weights pretrained on ImageNet. The model was trained with stochastic gradient descent algorithms in Keras with TensorFlow backend. As the different staining conditions could result in color variance between slides and have a negative impact on the training result, when training the model, image patches were augmented intensely before being fed into the model with self-defined image preprocessing functions. Each image had a 70% chance to be augmented. Augmentation methods included Gaussian blur, random projective transformations, color shifting for the whole image, color shifting and variation for each color channel, and random flipping horizontally and vertically.

### Slide-Level Prediction Heat Map Generation

A slide-level test data set was used to test the RMS classification model for whole slide prediction. This included the residual 192 slides not used in the model training procedures. All slides represented either aRMS ($n = 31$) or eRMS ($n = 161$), as too few scRMS slides were available, all of which were used in the training process. A 500 × 500 pixel window was then passed over the slides to generate image patches in the slide-level test data set with steps of 500 pixels. The image patches were predicted by the trained RMS subtype classification model. For each image patch, probabilities of being in each of the five tissue and region classes were generated, and the class with the highest probability was determined as the predicted class. Heat maps of the predicted class were generated for all the pathology image slides in the test data set.

### Slide-Level Image Classification

To generate the final prediction results of each slide, the above image patch prediction results were concluded by finding the RMS subtype class that constitutes the highest proportion of the predicted RMS tissue on the slide. Specifically, counts of aRMS, eRMS, and scRMS image patches from the prediction results were calculated for each slide in the slide-level test data set. The subtype class with the highest count was determined as the predicted RMS subtype result for the whole tissue slide. Receiver operating characteristic analysis was performed, and the measure used was area under the receiver operating characteristic curve.

## Transfer Learning for eRMS Patient Prognosis Prediction

### Transfer Learning Data Set Preparation

The survival rate for different subtypes of RMS is intrinsically different. To study the prognostic image features while eliminating the influence of subtype survival differences, eRMS, the largest subtype data set ($n = 196$) available, was

**Table 1** Patient Characteristics of the Whole Cohort Used in This Study

| Clinical characteristics | Value |
|---|---|
| Sex, *n* (%) | |
| Male | 185 (68.0) |
| Female | 87 (32.0) |
| Age at enrollment, days | |
| Median (range) | 2321 (22–13,803) |
| Race, *n* (%) | |
| Black or African American | 51 (18.8) |
| White | 192 (70.6) |
| Asian | 9 (3.3) |
| Native Hawaiian or other Pacific Islander | 1 (0.3) |
| Unknown or not reported | 19 (7.0) |
| Histology, *n* (%) | |
| aRMS | 66 (24.3) |
| eRMS | 196 (72.1) |
| scRMS | 10 (3.7) |
| T stage, *n* (%) | |
| 1: Favorable site | 86 (31.6) |
| 2: Unfavorable site, size ≤5 cm, and no regional node involvement | 29 (10.7) |
| 3: Unfavorable site, size >5 cm, and/or regional node involvement | 81 (29.8) |
| 4: Metastases | 76 (27.9) |
| Surgical group, *n* (%) | |
| I: Completely resected, margins and nodes negative | 33 (12.1) |
| II: Completely resected, margins positive | 29 (10.7) |
| III: Gross residual disease remains | 134 (49.3) |
| IV: Metastatic disease | 76 (27.9) |
| Pathologic characteristics | |
| Tumor size, median (range), cm | 5.6 (0.6–27) |
| *FOXO1* fusion status, *n* (%) | |
| *FOXO1*$^{+}$ | 26 (9.6) |
| Unknown or not reported | 246 (90.4) |
| Outcome characteristics | |
| Patient follow-up, median (range), days | 1830 (0–5321) |

aRMS, alveolar rhabdomyosarcoma; eRMS, embryonal rhabdomyosarcoma; scRMS, spindle cell sclerosing rhabdomyosarcoma.

chosen for transfer learning to establish an eRMS prognostic model. Slide images of eRMS patients with clear survival outcomes, either low risk (defined as patients with a >8-year progression-free period) or high risk (defined as patients with a disease progression during follow-up), were used to develop the patch-level prognostic model. Under these criteria, 30 low-risk and 30 high-risk eRMS slides were selected. These 60 slides were set to generate image patches to develop the model. Image patches with the size of 500 × 500 pixels under ×20 magnification were extracted from RMS tumor areas of these slides from expert annotation. Within the 30 low-risk patient slides, 24 were randomly chosen to generate 1337 training image patches, and six were used to generate 387 validation image patches; within the 30 high-risk patient-slides, 24 were randomly assigned and generated 1146 training image patches, and six were used to generate 204 validation image patches. The generated image patches in different data sets are summarized in Figure 2.

### Transfer Learning Process for Patch-Level eRMS Patient Prognosis Prediction

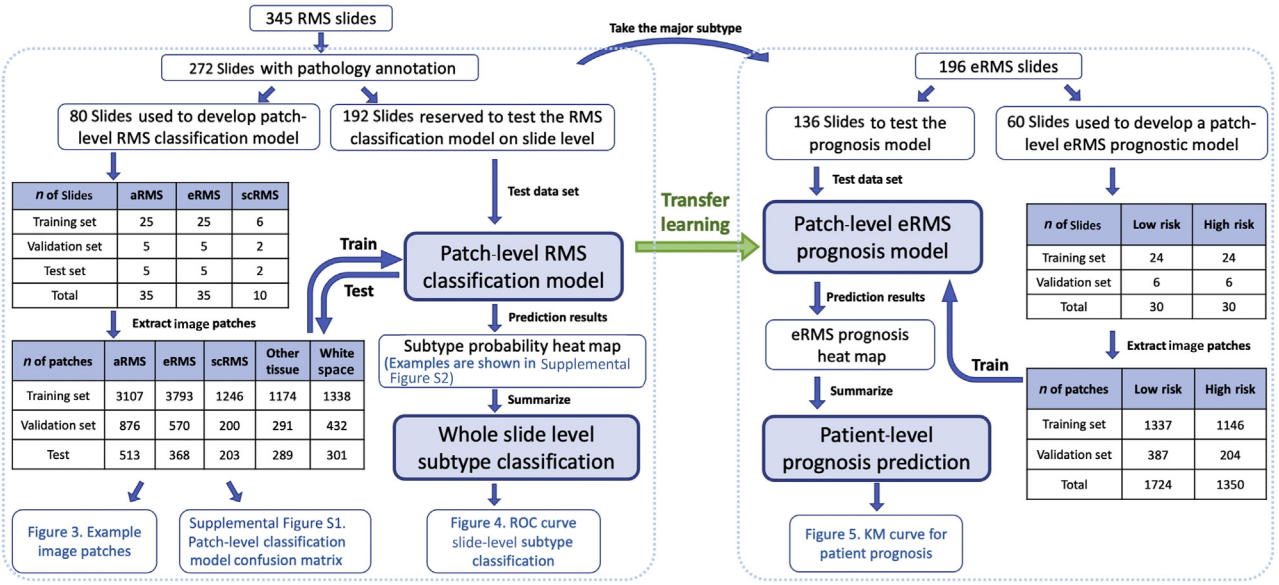
In this study, a patch-level eRMS patient prognosis prediction model was developed using transfer learning based on the previously trained RMS subtype classification model described in *Develop and Validate a Deep Learning Model for RMS Subtype Classification*. It adopts the same image augmentation method, which includes Gaussian blurring, color shifting, and projection changes of images. For the network architecture, the eRMS prognostic model used fixed layers before final flattening with fixed values from the RMS classification model, so the image features learned in the classification model can be utilized by the transfer learning model directly. Only the last four layers of flattening, dense, dropout, and another dense layer were retrained for eRMS prognosis model using new eRMS training image patches.

Data sets prepared in *Transfer Learning Data Set Preparation* were used to train the transfer learning eRMS prognosis prediction model. The model was also trained with stochastic gradient descent algorithms in Keras, and the training process was set to stop after 10 consecutive epochs of no validation accuracy improvement. The weights of the last training epoch were saved and were evaluated by predicting on the eRMS image patch test data set.

### Slide-Level Heat Map Generation of eRMS Prognosis Prediction

As mentioned in *Transfer Learning Data Set Preparation*, 60 slides were used to develop the image patch-level eRMS prognostic model. Among these 60 slides, 48 (24 low-risk patients and 24 high-risk patients) were used to generate training patches, and 12 (six low-risk patients and six high-risk patients) were used to generate validation patches. The remaining 136 eRMS slides (from the 196-slide eRMS cohort), which were not used in patch-level eRMS prognostic model development, were used to evaluate the performance of the eRMS prognostic model as the slide-level test data set for heat map generation and prognosis analysis.

The tissue image patches were predicted by the eRMS prognostic model. Only the image patches classified as RMS tumor areas by the RMS subtype classification model were included for the prognostic model, so that the other tissues and white areas were not used to predict patient prognosis. The probabilities of the RMS tumor image patches being from a low- or high-risk patient were predicted and recorded. A patch-level probability cutoff of being high risk was selected on the basis of the transfer learning model development data set ($n = 60$) for deciding




| *n* of Slides | aRMS | eRMS | scRMS |
|---|---|---|---|
| Training set | 25 | 25 | 6 |
| Validation set | 5 | 5 | 2 |
| Test set | 5 | 5 | 2 |
| Total | 35 | 35 | 10 |

| *n* of patches | aRMS | eRMS | scRMS | Other tissue | White space |
|---|---|---|---|---|---|
| Training set | 3107 | 3793 | 1246 | 1174 | 1338 |
| Validation set | 876 | 570 | 200 | 291 | 432 |
| Test | 513 | 368 | 203 | 289 | 301 |

| *n* of Slides | Low risk | High risk |
|---|---|---|
| Training set | 24 | 24 |
| Validation set | 6 | 6 |
| Total | 30 | 30 |

| *n* of patches | Low risk | High risk |
|---|---|---|
| Training set | 1337 | 1146 |
| Validation set | 387 | 204 |
| Total | 1724 | 1350 |

**Figure 2** A summary of data set, model development strategy (training, validation, and test), and results for this study. aRMS, alveolar RMS; eRMS, embryonal RMS; KM, Kaplan-Meier; RMS, rhabdomyosarcoma; ROC, receiver operating characteristic; scRMS, spindle cell sclerosing RMS.

if an RMS image patch in the heat map was from a low- or high-risk patient. For an individual patient, if most RMS image patches (>50%) from the slide had high risk based on the probability heat map, then the patient was predicted as high risk, and vice versa.

### Survival Analysis

Kaplan-Meier plots were used to summarize the event-free survival curves of the patients in predicted low- or high-risk groups. Likelihood ratio tests were performed to evaluate the survival difference between the two groups. A

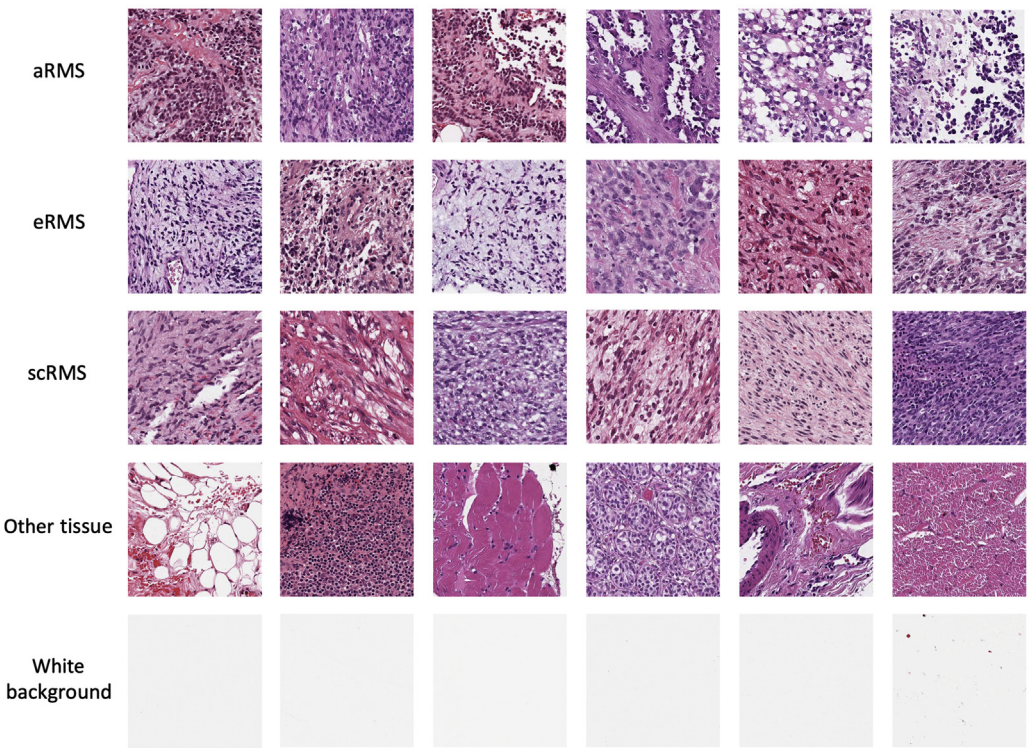



**Figure 3** Sample training patches from alveolar rhabdomyosarcoma (aRMS), embryonal rhabdomyosarcoma (eRMS), spindle cell sclerosing rhabdomyosarcoma (scRMS), other tissue (mainly connective tissue, but also including necrosis and glands), and white background categories. Patch size: 500 × 500 pixels. Original magnification, ×20.

multivariate Cox proportional hazard model was also used to analyze survival prediction adjusted for other clinical variables and to calculate the hazard ratio between the predicted high- versus low-risk group. All the survival analyses were performed with R software version 4.0.3 (The R Foundation, *https://www.r-project.org*). R package *survival* was used. The results were considered significant if the resulting two-tailed $P$ value was $<0.05$.

## Results

### RMS Subtype Classification Model

#### The Performance of RMS Subtype Classification Model at Image Patch Level

Eighty slides were used to generate 14,701 image patches for the development (including training, validation, and testing) of the RMS subtype classification model at image patch level. The image patches were separated into training, validation, and test data sets (see *Materials and Methods*). The model was trained on the training data set and simultaneously validated on the validation data set.

The trained RMS subtype classification model at image patch level was tested on the test data set composed of 1674 image patches (including 513 aRMS image patches, 368 eRMS image patches, 203 scRMS image patches, 289 image patches for other tissues, and 301 white background image patches). The resulting confusion matrix is shown in Supplemental Figure S1. The overall patch-level prediction accuracy of the five classes by the RMS subtype classification model in the test data set was 87.9%. The patch-level accuracy for the cancerous tissue was 84.0% for aRMS, 90.2% for eRMS, and 76.3% for scRMS. The accuracy for other (noncancerous) tissues was 87.5%; and for the white background, it was 99.7%.

#### Image Patch Prediction Heat Map Generation for the Whole Slide

After training, the established RMS classification model was used for slide-level subtype prediction. The 192 slides in the slide-level test data set were each partitioned into 500 × 500 image patches and predicted by the model to generate prediction heat maps showing the classification results. Some sample heat maps are shown in Supplemental Figure S2. For the original image in Supplemental Figure S2B, the prediction result in heat map (Supplemental Figure S2A) agreed with the pathologist's classification as aRMS; the slide (Supplemental Figure S2D) was classified by pathologist as eRMS, and the prediction result (Supplemental Figure S2C) showed the same judgment and classified most of the cancerous tissue as eRMS.

#### Slide-Level Subtype Classification

To summarize the image patch-level prediction results to the whole slide level (ie, for each patient), the patch-level predictions of the slides in the slide-level test data set were aggregated into slide classification results by finding the RMS subtype class that had the highest patch count on the heat map result. Of the 31 aRMS slides, 29 were correctly classified; 144 of 161 eRMS slides were also predicted accurately. In Figure 4, the receiver operating characteristic curves were plotted for the aRMS and eRMS slide-level prediction results using the calculated percentage of classes on each slide with the ground truth (which is the pathologist's diagnosis). The area under the curve values for aRMS and eRMS were 0.94 and 0.92, respectively.

### Prognosis Results on eRMS Patient Cohort

A total of 3074 image patches from low- or high-risk patient slides were used to train the transfer learning eRMS prognostic model. After eRMS prognostic model establishment, hematoxylin and eosin−stained tissue slide images from 136 eRMS patients (with one slide per patient) in the slide-level test data set were predicted by the model to generate prognosis prediction heat maps. A predefined cutoff value of the patch-level high-risk prediction was needed to determine the final slide-level classification. Herein, the cutoff was determined on the basis of the transfer learning model development data set and applied to the test data set ($n = 136$). After prediction, the 136 test slides were separated into 42 slides of low-risk prediction and 94 slides of high-risk prediction (Kaplan-Meier plot in Figure 5). The patients from the predicted low-risk group had significantly better survival with a much lower chance of having an event of relapse/progression, second malignancy, or death than the predicted high-risk group (likelihood ratio test; $P = 0.02$; Cox proportional hazards model predicted high- versus low-risk group hazard ratio, 4.55; 95% CI, 1.04−19.96; $P = 0.04$). Furthermore, Table 2 shows the multivariate analysis to adjust the image-based prognostic model with clinical variables using a Cox proportional hazard model. The results show that the image-based prognostic model predicts patient risk of relapse/progression (predicted high- versus low-risk group hazard ratio, 4.64; 95% CI, 1.05−20.57; $P = 0.04$) after adjusting for patient age and sex. The results indicate that the prognosis prediction model served as an independent prognostic factor for the patients. The established model has the potential to draw information from the slide images and contribute to the risk stratification of patients with pediatric eRMS.

## Discussion

Deep learning algorithms based on CNN for tumor region detection, subtype classification, and patient prognosis prediction for RMS histopathology images are presented here. The RMS classification model successfully recognized different tissue and background regions in the pathology images and correctly classified RMS slides into different histology subtypes. The eRMS prognostic model developed

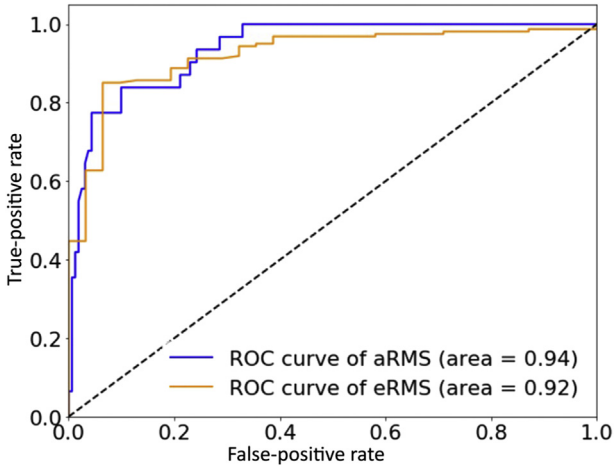



**Figure 4** Receiver operating characteristic (ROC) curve of alveolar rhabdomyosarcoma (aRMS) and embryonal rhabdomyosarcoma (eRMS) slide classification results.

by transfer learning successfully divided the eRMS patients into low- and high-risk groups, demonstrating the potential of extracting useful information from slide images for patient risk stratification. This finding could be useful in several contexts, such as to help physicians have a better idea of the risk for disease progression and to make better treatment recommendations.

The specific deep learning backbone selected here is the InceptionV3 CNN model. The InceptionV3 model features fewer parameters and higher computational efficiency.[29] The image patches were intensively augmented before being fed into the model to improve the model accuracy, which included Gaussian blurring, color shifting, and image

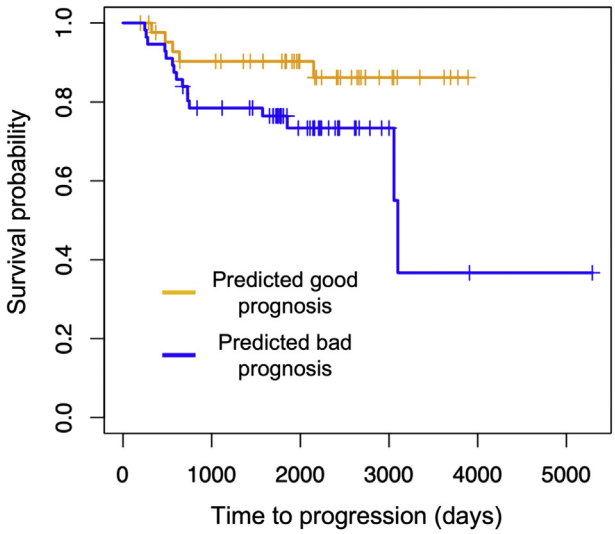



**Figure 5** Kaplan-Meier plot showing the prognostic performance of embryonal rhabdomyosarcoma transfer learning model. A total of 136 patients were dichotomized into 42 low risk and 94 high risk, according to the cutoff value of 0.30 for patch-level prediction in the slides. Difference between the two risk groups: likelihood ratio test, $P = 0.02$; Cox proportional-hazards model, $P = 0.04$.

**Table 2** Multivariate Analysis to Adjust Image-Based Prognostic Model with Clinical Variables Using Cox Proportional Hazard Model

| Clinical variables | HR | Lower 95% CI | Upper 95% CI | *P* value |
|---|---|---|---|---|
| Image-based prognostic model | 4.64 | 1.05 | 20.57 | 0.04 |
| Age | 1.00 | 1.00 | 1.00 | 0.78 |
| Sex (male) | 1.04 | 0.37 | 2.93 | 0.94 |

HR, hazard ratio.

projection. The RMS classification model was trained on a training data set of 10,658 image patches and tested on a patch-level test data set of 1674 image patches, showing a test accuracy of 87.9%. Applying the RMS classification model for prediction on the slide-level testing data set showed an area under the receiver operating characteristic curve of 0.94 for aRMS slides and 0.92 for eRMS slides. These results point to the successful application of CNN for the rare disease RMS subtype classification. Transferring the information learned in the classification model for an eRMS prognostic model for low- and high-risk patient stratification resulted in a hazard ratio of 4.64 ($P = 0.04$) with the Cox proportional hazard model after adjusting for patient age and sex.

Before this work, little effort in machine learning application was made for this rare disease. Only studies reporting deep learning for RMS with other imaging modalities, like magnetic resonance imaging, were available.[30,31] This is one of the first studies[28] to apply deep learning on histopathology images of RMS. As correctly classifying different subtypes of RMS requires the special training and expertise of a pathologist, this tool might provide support for such expertise.

There are still limitations to this study. First, because of their rarity, the imaging data for some patterns of RMS were limited and not sufficient to establish a robust deep learning model, and were thus excluded in this study. In addition to the typical patterns of the three main RMS subtypes in this study, there can be anaplasia variants of these RMS subtypes, or sclerosing variants of RMS. These are real-world scenarios that were not considered in this study because of the limitations of the image data. Apart from this, only one specific network of choice, InceptionV3, was applied based on previous experience, without comparing more commonly used CNN models. Although satisfying results were already achieved, the model can be improved by testing on more network structures and building ensemble models from the best performing networks.[32]

There are several directions for future extension and improvement of the current study. The first is to continue improving our RMS classification and prognostic models, by refining the CNN architecture and including more available slide images for training, to further improve the model performances. The prognostic model can also be expanded from eRMS to other RMS subtypes when there

are more image data available. Furthermore, integrating different dimensions of patient data, such as genomic data, to the current framework is another key goal. Combining genetic information with image feature information, a better prognostic model may be established for risk prediction of RMS patients.

## Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.ajpath.2022.03.011*.

## References


1. Ries LAG: Cancer Incidence and Survival Among Children and Adolescents: United States SEER Program, 1975-1995. Bethesda, MD, National Cancer Institute, 1999
2. Newton WA Jr, Gehan EA, Webber BL, Marsden HB, van Unnik AJM, Hamoudi AB, Tsokos MC, Shimada H, Harms D, Schmidt D, Ninfo V, Cavazzana AO, Gonzalez-Crussi F, Parham DM, Reiman HM, Asmar L, Beltangady MS, Sachs NE, Triche TJ, Maurer HM: Classification of rhabdomyosarcomas and related sarcomas: pathologic aspects and proposal for a new classification–an Intergroup Rhabdomyosarcoma Study. Cancer 1995, 76:1073−1085
3. Bridge J, Hogendoorn P, DM C, Bridge JA, CW P, Fletcher CD: WHO Classification of Tumours of Soft Tissue and Bone. Lyon, France, International Agency for Research on Cancer, 2013
4. Rudzinski ER, Kelsey A, Vokuhl C, Linardic CM, Shipley J, Hettmer S, Koscielniak E, Hawkins DS, Bisogno G: Pathology of childhood rhabdomyosarcoma: a consensus opinion document from the Children's Oncology Group, European Paediatric Soft Tissue Sarcoma Study Group, and the Cooperative Weichteilsarkom Studiengruppe. Pediatr Blood Cancer 2021, 68:e28798
5. Tsokos M, Webber B, Parham D, Wesley R, Miser A, Miser J, Etcubanas E, Kinsella T, Grayson J, Glatstein E: Rhabdomyosarcoma: a new classification scheme related to prognosis. Arch Pathol Lab Med 1992, 116:847−855
6. Kodet R, Newton WA Jr, Hamoudi AB, Asmar L, Wharam MD, Maurer HM: Orbital rhabdomyosarcomas and related tumors in childhood: relationship of morphology to prognosis—an Intergroup Rhabdomyosarcoma Study. Med Pediatr Oncol 1997, 29:51−60
7. Crist WM, Anderson JR, Meza JL, Fryer C, Raney RB, Ruymann FB, Breneman J, Qualman SJ, Wiener E, Wharam M: Intergroup rhabdomyosarcoma study-IV: results for patients with nonmetastatic disease. J Clin Oncol 2001, 19:3091−3102
8. Sorensen PH, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, Bridge JA, Crist WM, Triche TJ, Barr FG: PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the Children's Oncology Group. J Clin Oncol 2002, 20:2672−2679
9. Williamson D, Missiaglia E, de Reynies A, Pierron G, Thuille B, Palenzuela G, Thway K, Orbach D, Laé M, Fréneaux P: Fusion gene-negative alveolar rhabdomyosarcoma is clinically and molecularly indistinguishable from embryonal rhabdomyosarcoma. J Clin Oncol 2010, 28:2151−2158
10. Skapek SX, Anderson J, Barr FG, Bridge JA, Gastier-Foster JM, Parham DM, Rudzinski ER, Triche T, Hawkins DS: PAX-FOXO1 fusion status drives unfavorable outcome for children with rhabdomyosarcoma: a Children's Oncology Group report. Pediatr Blood Cancer 2013, 60:1411−1417
11. Shenoy A, Alvarez E, Chi Y-Y, Li M, Shern JF, Khan J, Hiniker SM, Granberg CF, Hawkins DS, Parham DM: The prognostic significance of anaplasia in childhood rhabdomyosarcoma: a report from the Children's Oncology Group. Eur J Cancer 2021, 143:127−133
12. Wang S, Yang DM, Rong R, Zhan X, Xiao G: Pathology image analysis using segmentation deep learning algorithms. Am J Pathol 2019, 189:1686−1698
13. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 2015, 521: 436−444
14. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, Moreira AL, Razavian N, Tsirigos A: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018, 24:1559−1567
15. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A, Campilho A: Classification of breast cancer histology images using convolutional neural networks. PLoS One 2017, 12:e0177544
16. Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012, 25:1097−1105
17. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017, 318:2199−2210
18. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, Thng F, Peng L, Stumpe MC: Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. Am J Surg Pathol 2018, 42:1636−1646
19. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, Venugopalan S, Timofeev A, Nelson PQ, Corrado GS: Detecting cancer metastases on gigapixel pathology images. arXiv 2017. [Preprint] doi:10.48550/arXiv.1703.02442
20. Dalle J-R, Leow WK, Racoceanu D, Tutac AE, Putti TC: Automatic breast cancer grading of histopathological images. IEEE, 2008. pp. 3052−3055. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society
21. Veta M, Van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, Gonzalez F, Larsen AB, Vestergaard JS, Dahl AB: Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med Image Anal 2015, 20:237−248
22. Albawi S, Mohammed TA, Al-Zawi S: Understanding of a convolutional neural network. IEEE, 2017. pp. 1−6. 2017 International Conference on Engineering and Technology (ICET), Aug. 21-23, 2017
23. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI: A survey on deep learning in medical image analysis. Med Image Anal 2017, 42:60−88
24. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q: A comprehensive survey on transfer learning. Proc IEEE 2020, 109: 43−76
25. Rai T, Morisi A, Bacci B, Bacon NJ, Thomas SA, La Ragione RM, Bober M, Wells K: An investigation of aggregated transfer learning for classification in digital pathology Proc. SPIE 10956, Medical Imaging 2019: Digital Pathology; 2019. 109560U
26. Xu Y, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, Chang EI: Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinformatics 2017, 18:281
27. Kieffer B, Babaie M, Kalra S, Tizhoosh HR: Convolutional neural networks for histopathology image classification: training vs. using pre-trained networks. IEEE, 2017. pp. 1−6. 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)
28. Agarwal S, Eltigani M, Abaker O, Zhang X, Daescu O, Barkauskas DA, Rudzinski ER, Leavey P: Rhabdomyosarcoma histology classification using ensemble of deep learning networks. Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Association for Computing Machinery: Virtual Event, USA: ACM, 2020. Article 24
29. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z: Rethinking the inception architecture for computer vision. IEEE, 2016. pp.

2818−2826. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

30. Banerjee I, Crawley A, Bhethanabotla M, Daldrup-Link HE, Rubin DL: Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput Med Imaging Graph 2018, 65:167−175
31. Hermessi H, Mourali O, Zagrouba E: Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning. Expert Syst Appl 2019, 120:116−127
32. Perez F, Avila S, Valle E: Solo or ensemble? Choosing a CNN architecture for melanoma classification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2019