Predicting Femoral Head Collapse Risk in Osteonecrosis Using Label Tokenization: A Multi-Modality Survival Analysis Approach

Ganping Li[1], Yoshito Otake[1], Yuito Kameda[1], Keisuke Uemura[2], Kazuma Takashima[2], Hirokazu Mae[2], Sotaro Kono[2], Hidetoshi Hamada[2], Seiji Okada[2], Nobuhiko Sugano[2], and Yoshinobu Sato[1⋆]

[1] Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan
{li.ganping.lc2,otake,yoshi}@naist.ac.jp
[2] Department of Orthopaedics, Osaka University Graduate School of Medicine, Japan

Abstract. Collapse of the femoral head is a critical event in osteonecrosis (ONFH) that often leads to debilitating hip pain and necessitates total hip arthroplasty. Early and accurate prediction of collapse risk is crucial for personalized treatment planning. While many studies focus on the automated diagnosis of ONFH, prognosis remains less explored. In this study, we propose a robust tri-stream deep learning framework that extracts features from T1-weighted MRI, region-of-interest (ROI) labels, and ONFH grades to estimate patient-specific collapse risk. We introduce an independent Spatial Label Encoder (SLE) module that tokenizes discrete ROI labels into dense, context-rich embeddings, thereby facilitating multi-modality model training. Experiments on 92 hips (70 patients) show that our approach performs competitively with state-of-the-art (SOTA) methods across most metrics, achieving a concordance index (CI) of 0.847±0.087 and an integrated AUC of 0.884 in 5-fold cross-validation. Notably, the SLE module enhances long-term discrimination by up to 2.4% on AUC at 60 months compared to our base network. These findings highlight the potential benefits of late-fusion strategies with label tokenization for predicting femoral head collapse in ONFH, contributing to improved early intervention and prognosis.

Keywords: Survival Prediction · Osteonecrosis of the Femoral Head · Representation Learning · Deep Survival Model.

1 Introduction

Osteonecrosis of the femoral head (ONFH) is a major health concern among young and middle-aged individuals, often leading to severe hip pain and dysfunction as it progresses toward femoral head collapse [2, 5]. Since collapse is a

⋆ Currently at Department of Mathematical and AI Medical Science, Nara Medical University.

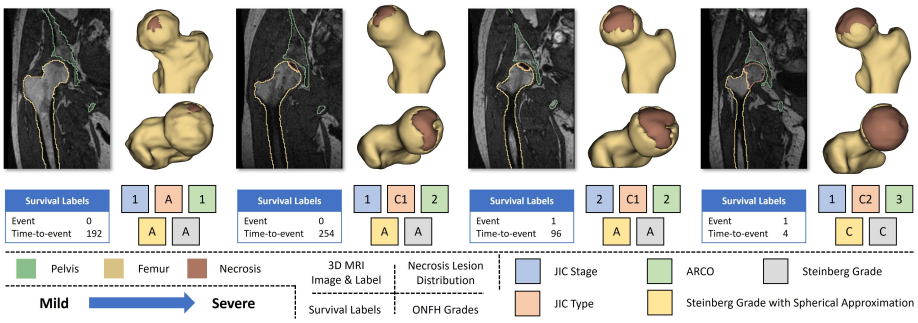

Fig. 1. Representative samples from our ONFH dataset, each annotated with multiple grade labels from different grading systems due to the lack of a universal standard. The severity of ONFH increases from left to right across the four samples.

pivotal turning point in ONFH, timely risk assessment is crucial for determining whether patients may benefit from conservative treatment or require immediate hip preservation surgery, such as total hip arthroplasty (THA) [8, 12]. Figure 1 illustrates representative samples of ONFH severity progression, central to our study on femoral head collapse prediction. While many recent studies focus on diagnosing ONFH, the risk of subsequent femoral head collapse is often overlooked [11, 20]. Although manual risk assessment methods have been proposed for prognostic studies of ONFH, they are time-consuming and prone to observer variability. As a result, automated deep learning approaches have gained attention for their efficiency and consistency [6].

A key challenge in predicting ONFH femoral head collapse risk is accurately modeling the relationship between the risk score and its covariates (e.g., patient data). Conventional survival prediction often relies on predefined Radiomics-based features derived from segmented regions of interest (ROI) [26], which is then processed by either statistical models (e.g., Cox Proportional Hazard (CoxPH) [3]) or machine learning models (e.g., survival support vector machine (SVM) [25]). Recent studies have proposed deep learning-based survival models with fusion mechanisms to better capture complex associations between covariates and survival outcomes. These methods employ various convolutional neural networks (CNNs)- or transformer-based encoders to leverage latent information from multimodal data such as radiological scans, electronic health records (EHR), and whole slide images (WSI) [21, 22, 27].

However, the presence of limited follow-up data (i.e., right-censored data) poses a challenge to the performance of survival prediction models. To address this, Qu et al. [19] introduced an auxiliary network to estimate hazards for censored intervals by leveraging statistics from uncensored and extended follow-up samples. In contrast, SurvRNC [22] captures inter- and intra-class relationships among censored and uncensored samples through representation alignments in survival-time order. Despite these advancements, both approaches rely on early-fusion strategies (e.g., concatenating multimodal images or spatial segmentations

at the input stage) to integrate multi-modality or region-specific features. While recent studies highlight the importance of region-specific information, such as the joint learning of segmentation and prognosis tasks [15] or the early fusion of ROI labels and images [24], few have explored late-fusion techniques that incorporate tokenized semantic label features to enhance modality alignment.

In this paper, we propose a robust framework for predicting femoral head collapse in ONFH patients by extracting latent representations from pre-collapse MRI scans, segmentation masks, and tabular grades using a tri-stream encoder. These representations are fused via a multilayer perceptron (MLP). In addition, a lightweight Spatial Label Encoder (SLE) that parses discrete semantic labels into learnable embeddings is proposed. This design yields compact, context-rich representations of label distributions that can be seamlessly integrated with image-based features for survival analysis.

The contributions of this paper are three-fold: 1) to the best of our knowledge, this is the first general survival analysis framework for multi-modality femoral head collapse risk prediction, which could facilitate personalized surgical planning (e.g., THA) for ONFH patients; 2) the impact of incorporating the proposed SLE module is assessed on both our model and a state-of-the-art (SOTA) deep learning model, showing consistent performance gains; 3) the effectiveness of the proposed framework is demonstrated and compared with several SOTA methods using a 92-case MRI dataset of pre-collapse ONFH hips.

2 Method

2.1 Dataset and Preprocessing

We retrospectively collected T1-weighted MRI scans from the first hospital visit of 70 patients with ONFH, covering 92 hips (Japanese Investigation Committee (JIC) classification: 52 hips in stage 1, 40 in stage 2), as illustrated in Figure 1. All MRI scans were bisected and flipped to the right for left hip targets, resulting in volumes of shape $(128, 256, n)$ with voxel spacing $(1.25, 1.25, 1)$ mm, where $n \in [54, 159]$. Among these, 63 hips were manually annotated by an experienced orthopedic surgeon, while the remaining 29 were automatically segmented using a Dynamic U-Net [9, 1] trained on the labeled cases. The mean Dice coefficients for the pelvis, femur, and necrosis were 0.931, 0.945, and 0.890, respectively. Since no universal ONFH grading standard exists, each hip was assigned multiple classifications: JIC stage, JIC type, Association Research Circulation Osseous classification (ARCO), Spherical Approximation Steinberg grade, and Steinberg classification [23]. To address the modest dataset size and ensure robust findings, all experiments employed a strict patient-level 5-fold cross-validation. We are also working to secure permissions to release the data publicly in the future.

Femoral head collapse during the study period was considered the event of interest, with time-to-event measured from the first visit to either the occurrence of collapse (uncensored data) or the last available follow-up (right-censored). Approximately 47% of patients had uncensored data, with time-to-event ranging

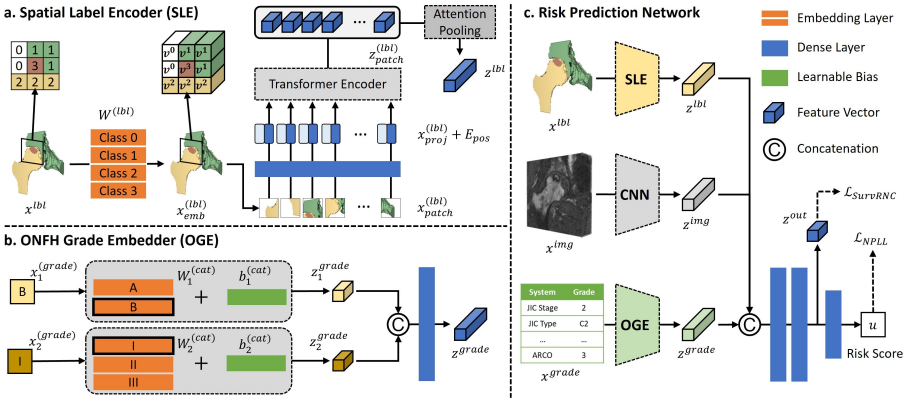

Fig. 2. Framework overview: a. **Spatial Label Encoder** maps discrete class labels to n-dim learnable vectors and extracts spatial and inter-class features in a ViT-like manner. **b. ONFH Grade Embedder** tokenizes categorical grading inputs and projects the tokens into grade features. **c. Risk Prediction Network** predicts the risk of femoral head collapse using the patient's MRI scans, ROI labels, and ONFH grades.

from 1 to 174 months. The median time-to-event was 12 months (IQR: 3.5–33 months), with 90% of events occurring within 82 months.

2.2 Femoral Head Collapse Risk Prediction

Problem Statement This task aims to predict the risk of femoral head collapse in ONFH hips using multimodal data: MRI scans x_{img} of size (H, W, D), ROI labels x_{lbl}, and ONFH grades x_{grade}. We denote our dataset by $\mathcal{D} = \{\mathcal{H}_1, \ldots, \mathcal{H}_N\}$, where each hip $\mathcal{H}_i$ comprises features $X_i = (x_i^{img}, x_i^{lbl}, x_i^{grade})$, an event indicator e_i (0 for censored, 1 for uncensored), and a time-to-event T_i. The objective is to train a deep network f to predict a risk score $u = f(X)$ (with higher values indicating a worse prognosis), which can be further interpreted by survival functions $S(t \mid X)$ (e.g., via the Kaplan-Meier estimator [10]).

Spatial Label Encoder Given a 3D image and its ROI label, a standard approach is to concatenate the scalar or one-hot label with the image to get $\mathbf{X} = [x^{img}; x^{lbl}]$ in an early fusion mechanism. While intuitive and straightforward, this approach has two drawbacks: 1) the discrete and sparse nature of categorical labels introduces redundancy, especially in multi-class settings [29, 28]; and 2) early fusion can lead to insufficient intra-modality information due to entangled (concatenated) features [15]. To address these issues, we propose SLE that tokenizes the discrete multi-class label map and encodes it with a shallow vision transformer (ViT) [4], as illustrated in Figure 2a. Given a scalar-valued label map $x^{lbl} \in \mathbb{R}^{1\times H\times W\times D}$, an embedding matrix $W^{(lbl)} \in \mathbb{R}^{C\times d}$ maps each class index c to a d-dimensional vector $v^{cls} = W^{(lbl)}[c]$, yielding a rich, learnable

representation compared to sparse one-hot labels. The embedded labels $x_{emb}^{(lbl)}$ are then passed through a shallow ViT, and the attention pooling filters out low-information patches, producing a final label feature $z^{lbl} \in \mathbb{R}^{d^{lbl}}$ for downstream fusion:

$$x_{emb}^{(lbl)} = W^{(lbl)}(x^{lbl}) \in \mathbb{R}^{d \times H \times W \times D} \tag{1}$$

$$z^{lbl} = \mathrm{ViT}(\mathrm{PatchEmbed}(x_{emb}^{(lbl)})) \in \mathbb{R}^{d^{lbl}} \tag{2}$$

The class tokenization supports the established idea that high-dimensional, continuous features enhance a model's ability to capture subtle variations [17, 13], while the ViT-like module provides a spatially rich representation of label distributions. Since the SLE is an independent module, it supports late-fusion frameworks and can be seamlessly integrated into other SOTA methods.

ONFH Grade Embedder This module follows the same scalar-value tokenization concept in SLE and leverages the Feature Tokenizer Transformer (FT-Transformer) [7] for categorical value embedding, as illustrated in Figure 2.b. Given a categorical index x_j^{grade}, where $j \in \{1, 2, \ldots, G\}$ (and G is the number of grade systems), we define:

$$z_j^{(grade)} = b_j^{(cat)} + W_j^{(cat)}\left(x_j^{grade}\right) \quad \in \mathbb{R}^d, \tag{3}$$

$$z^{grade} = f_{\mathrm{proj}}\left([z_1^{(grade)}; z_2^{(grade)}; \ldots; z_G^{(grade)}]\right) \in \mathbb{R}^{d^{grade}}, \tag{4}$$

where $b_j^{(cat)}$ represents the bias feature in the j-th categorical system, f_{proj} denotes a linear projection, and z^{grade} is the embedded grading feature of dimension d^{grade}.

Proposed Framework Figure 2.c demonstrates the overall risk prediction network. A tri-stream encoder, consisting of CNN-based vision encoder (shallow ResNet-18) f_{CNN}, SLE f_{SLE}, and ONFH grade embedder f_{OGE}, takes input $X = [x^{img}, x^{lbl}, x^{grade}]$ and outputs $z^{img} \in \mathbb{R}^{d^{img}}$, z^{lbl}, and z^{grade}, respectively. The three feature vectors are then fused by simple concatenation and passed to a 2-layer MLP to get feature representations z^{out}, and the final output risk score u is predicted from z^{out} through a dense output layer. To optimize model parameters, we combine the negative partial log-likelihood loss $\mathcal{L}_{\mathrm{NPLL}}(u, e, T)$ [3] with an ordered representation alignment loss $\mathcal{L}_{\mathrm{SurvRNC}}(\cdot)$ [22], where anchor, positive, and negative embeddings $(z_a^{out}, z_p^{out}, \{z_k^{out}\}_{k \in \mathcal{N}_{a,p}})$ enforce ordinal constraints. The total loss is:$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{NPLL}} + \lambda * \mathcal{L}_{\mathrm{SurvRNC}}$, with λ empirically set to 0.5.

2.3 Evaluation Metrics

Four metrics were used to thoroughly evaluate our method: the concordance index (CI), the time-dependent Area Under the Curve (AUC(t)), the Integrated

Table 1. Performance of different SOTA survival prediction methods on a 5-fold cross-validation evaluation. The best prediction values are shown in bold. Note that *CoxPH was trained only by ONFH grades.

Metrics →	**CI** ↑	**AUC**(t) ↑					**iAUC** ↑
Methods↓		6(m)	12(m)	36(m)	60(m)	120(m)	12∼120(m)
CoxPH* [3]	0.820±0.059	0.875±0.061	0.869±0.082	0.872±0.070	0.853±0.108	0.852±0.108	0.860±0.096
SurvivalSVM [25]	0.781±0.095	0.829±0.160	0.856±0.168	0.882±0.106	0.858±0.073	0.806±0.103	0.851±0.081
DeepMTS [16]	0.824±0.093	0.825±0.107	0.841±0.111	0.857±0.075	0.867±0.069	0.840±0.111	0.861±0.078
XSurv [15]	0.822±0.068	0.878±0.080	**0.912±0.064**	**0.920±0.065**	0.856±0.079	0.783±0.139	0.865±0.081
SurvRNC [22]	0.832±0.094	0.877±0.086	0.878±0.122	0.863±0.078	0.868±0.104	0.858±0.132	0.869±0.103
Ours (w/o SLE)	0.838±0.076	0.879±0.101	0.885±0.136	0.874±0.077	0.853±0.087	0.866±0.117	0.866±0.091
SurvRNC (w SLE)	0.841±0.088	0.879±0.096	0.894±0.120	0.875±0.071	0.868±0.102	0.860±0.120	0.873±0.098
Ours (w SLE)	**0.847±0.087**	**0.883±0.105**	0.885±0.122	0.880±0.068	**0.877±0.085**	**0.883±0.106**	**0.884±0.085**

AUC (iAUC), and KM analysis. CI, as the primary metric, measures the model's overall discriminative power (ranging from 0 to 1). In contrast, AUC(t) evaluates the model's time-specific discrimination ability to distinguish individuals who experience an event before time t versus those surviving beyond t. iAUC is the aggregated version of AUC(t) that reflects overall discrimination ability over a time range $[t_{min}, t_{max}]$. Finally, KM analysis determines the stratification efficacy of the model by dividing patients into high-risk and low-risk groups.

3 Experiments and Results

3.1 Baselines and Configurations

To evaluate the validity of our proposed method, we conducted comparative experiments with several SOTA multi-modality networks that use 3D medical images as their primary input. Specifically, we assessed DeepMTS [16], XSurv [15], and SurvRNC [22] using a 5-fold patient-wise cross-validation on our femoral head collapse dataset. During training, 20% of the hip joints were set aside as the validation set for each fold. For a fair comparison, we replaced the original models' sub-modalities and clinical indicators with comparable sources of collapse-related information (as used by [6] for ROI labels and [2] for ONFH grades). Additionally, we compared our method with two conventional survival models, CoxPH and SurvivalSVM, using Radiomics features to represent both image and label modalities. Radiomics feature extraction and selection followed the preprocessing pipeline described in [15, 6]. For all methods (both classical and deep learning), we report only the highest evaluation scores across different modality combinations.

All deep learning models were fine-tuned based on their default hyperparameters and survival losses ($\mathcal{L}_{\text{NPLL}}$ or $\mathcal{L}_{\text{Total}}$), trained and tested on a computing device with an Intel Xeon W-2295 @3.00GHz CPU, 125GB RAM, and an NVIDIA A6000 GPU (48 GB) using PyTorch [18] version 2.0.1 with CUDA 11.8. Each model underwent the same data augmentations and was trained for 310 epochs with a batch size of 16, a learning rate of 4×10^{-5}, and the AdamW optimizer [14] (weight decay of 10^{-4}).

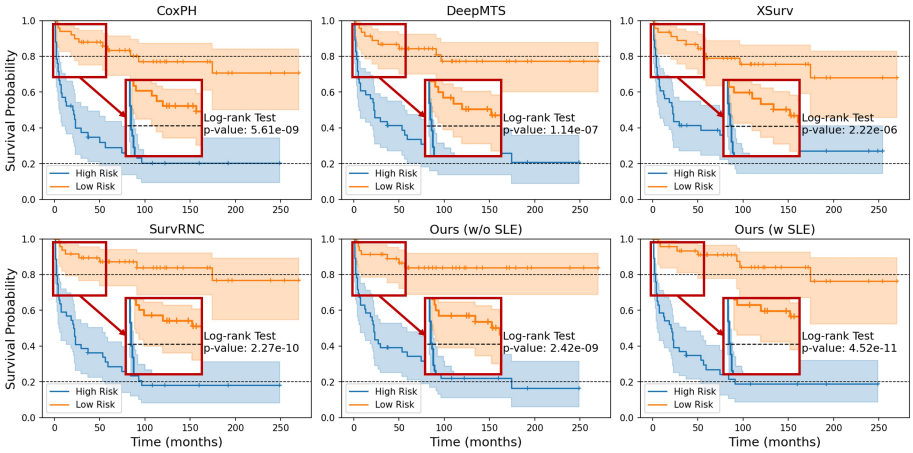

Fig. 3. KM analysis for CoxPH and deep learning methods on our ONFH dataset. Our full method with SLE achieves the best p-value in the log-rank test while our base network demonstrates the best stratification ability at the end of the censor duration.

To further validate our approach, we tested our method with the SLE module replaced by early fusion to confirm the effectiveness of the base network. Additionally, we integrated our SLE module into SurvRNC to demonstrate its adaptability and efficacy. For our proposed method, the dimensions d^{img}, d^{lbl}, and d^{grade} were empirically set to 128, 128, and 64, respectively.

3.2 Results and Discussions

Table 1 compares the performance of all methods across three metrics. Notably, CoxPH achieves its best performance when trained solely on categorical ONFH grades, whereas other methods require all modalities. Our proposed base network (w/o SLE) achieves the highest CI (0.838) among previous methods and demonstrates strong short- to mid-term AUC ($t \leq 36$), highlighting the effectiveness of tokenization-based embedding for tabular data. When integrated with SLE, our method addresses the long-term discrimination issue ($t = 60, 120$) and shows improvements over most metrics (except mid-term $t = 12, 36$). In particular, AUC(60) increases by 0.024, AUC(120) by 0.017, and iAUC by 0.018. Adding SLE to the second-best (w/o SLE) method SurvRNC also improves all metrics, suggesting enhanced feature representation while maintaining model stability.

Regarding baseline methods, CoxPH serves as a strong competitor when trained only on categorical ONFH grades, aligning with previous studies [22, 15] given the close relationship between disease grades and ONFH prognosis [6]. In contrast, SurvivalSVM performs the worst on CI, likely due to convergence failure by the limited amount of trainable data. Among deep learning approaches, DeepMTS slightly outperforms CoxPH on CI but performs worse in AUC, possibly due to its entangled volume inputs and its original multitask design. XSurv

achieves the highest mid-term AUC(36) at 0.92, which may benefit from its dual symmetric encoder of the same perception field for volumes from two modalities with aligned spatial information [27]. However, the deeper network is more sensitive to a small, right-skewed dataset (as is common with uncensored data in medical scenarios), resulting in poorer long-term AUC. SurvRNC, despite employing early fusion, provides balanced results across different time windows thanks to its effective representation alignment loss $\mathcal{L}_{\text{SurvRNC}}$. Our proposed method integrates key multimodal mechanisms from previous SOTA approaches while addressing their limitations by: 1) disentangling spatial modalities via multiple lightweight encoders for late fusion, 2) facilitating implicit cross-modality alignment by maintaining tokenizers in the SLE and ONFH grade encoders, and 3) aligning representation in survival-time order through $\mathcal{L}_{\text{SurvRNC}}$.

To further assess stratification efficacy, Fig. 3 presents KM curves for CoxPH and various deep learning methods, with shaded regions indicating confidence intervals. The log-rank test p-values confirm statistically significant differences between risk groups. While our base network shows a larger discrepancy at the final time ($t > 174$), incorporating SLE results in a smaller p-value between risk groups. Our full method with SLE also yields a more stable survival curve for $t < 60$ in the low-risk group and a steeper early decline in the high-risk group, leading to a more distinct separation between risk groups earlier, which is clinically valuable for predicting femoral head collapse where timely intervention is critical. These findings align with the numerical estimates in Table 1, demonstrating a moderate overall advantage of our approach over other SOTA methods on the ONFH dataset, as reflected by lower p-values and earlier curve separation.

3.3 Limitations and Future Works

The proposed framework has two main limitations. First, our study is based on a small, single-center retrospective dataset, which may impact the generalizability of the evaluated methods. Some ground-truth labels may also contain noise due to reliance on automated segmentation. Second, all modalities in our experiments are pre-aligned temporally, and volume-based modalities (MRI and its embedded ROI labels) are also spatially aligned in advance. The impact of these alignments requires further investigation, particularly in the context of follow-up data integration of different modalities and positioning.

Future work will focus on enhancing robustness by incorporating external datasets and mitigating label noise through uncertainty-aware modeling. We also plan to integrate additional imaging modalities with diverse spatial and temporal characteristics (e.g., X-ray and follow-up Computed Tomography) to further assess the model's generalizability. Moreover, a comprehensive analysis of each modality's contribution is needed. Further refinements to the framework will aim to maximize the benefits of SLE. Lastly, ongoing follow-up data collection will support more comprehensive comparisons in future studies.

4 Conclusions

This study presents a tri-stream framework[1] that integrates MRI scans, ROI labels, and ONFH grades to predict femoral head collapse risk, while also evaluating several SOTA multi-modality methods in this context. By combining label tokenization and time-ordered representation alignments, our approach reduces the redundancy of scalar or one-hot multi-class labels while facilitating implicit cross-modality alignment. The proposed SLE has demonstrated promising and consistent improvements, suggesting its potential as an enhanced label-encoding strategy. While further validation is required, our framework provides a structured and robust baseline for ONFH prognosis, with the potential to aid early clinical interventions and advance research in personalized medicine.

Acknowledgments. This study was funded by MEXT/JSPS KAKENHI (19H01176, 20H04550, 21K16655, 23K15714) and AMED under Grant Number JP25hma322015.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
2. Chen, L., Hong, G., Fang, B., Zhou, G., Han, X., Guan, T., He, W.: Predicting the collapse of the femoral head due to osteonecrosis: from basic methods to application prospects. Journal of Orthopaedic Translation **11**, 62–72 (2017)
3. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
5. Fan, Y., Liu, X., Zhong, Y., Zhang, J., Liu, Y., Fang, H., He, W., Zhou, C., Chen, Z.: Evaluation of the predictive values of collapse and necrotic lesion boundary for osteonecrosis of the femoral head prognosis. Frontiers in Endocrinology **14**, 1137786 (2023)
6. Gao, S., Zhu, H., Wen, M., He, W., Wu, Y., Li, Z., Peng, J.: Prediction of femoral head collapse in osteonecrosis using deep learning segmentation and radiomics texture analysis of mri. BMC Medical Informatics and Decision Making **24**(1), 320 (2024)
7. Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A.: Revisiting deep learning models for tabular data. Advances in neural information processing systems **34**, 18932–18943 (2021)
8. Hindoyan, K.N., Lieberman, J.R., Matcuk Jr, G.R., White, E.A.: A precise and reliable method of determining lesion size in osteonecrosis of the femoral head using volumes. The Journal of Arthroplasty **35**(1), 285–290 (2020)

[1] The source code is available at: https://github.com/RIO98/FemoralCollapsePrediction

9. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
10. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American statistical association **53**(282), 457–481 (1958)
11. Klontzas, M.E., Vassalou, E.E., Spanakis, K., Meurer, F., Woertler, K., Zibis, A., Marias, K., Karantanas, A.H.: Deep learning enables the differentiation between early and late stages of hip avascular necrosis. European Radiology **34**(2), 1179–1186 (2024)
12. Kuroda, Y., Tanaka, T., Miyagawa, T., Kawai, T., Goto, K., Tanaka, S., Matsuda, S., Akiyama, H.: Classification of osteonecrosis of the femoral head: who should have surgery? Bone & joint research **8**(10), 451–458 (2019)
13. Li, G., Otake, Y., Soufi, M., Masuda, M., Uemura, K., Takao, M., Sugano, N., Sato, Y.: Prediction of disease-related femur shape changes using geometric encoding and clinical context on a hip disease ct database. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 368–378. Springer (2024)
14. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
15. Meng, M., Bi, L., Fulham, M., Feng, D., Kim, J.: Merging-diverging hybrid transformer networks for survival prediction in head and neck cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 400–410. Springer (2023)
16. Meng, M., Gu, B., Bi, L., Song, S., Feng, D.D., Kim, J.: Deepmts: Deep multitask learning for survival prediction in patients with advanced nasopharyngeal carcinoma using pretreatment pet/ct. IEEE Journal of Biomedical and Health Informatics **26**(9), 4497–4507 (2022)
17. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
18. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
19. Qu, L., Huang, D., Zhang, S., Wang, X.: Multi-modal data binding for survival analysis modeling with incomplete data and annotations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 501–510. Springer (2024)
20. Rakhshankhah, N., Abbaszadeh, M., Kazemi, A., Rezaei, S.S., Roozpeykar, S., Arabfard, M.: Deep learning approach to femoral avn detection in digital radiography: differentiating patients and pre-collapse stages. BMC Musculoskeletal Disorders **25**(1), 547 (2024)
21. Ramanathan, V., Pati, P., McNeil, M., Martel, A.L.: Ensemble of prior-guided expert graph models for survival prediction in digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 262–272. Springer (2024)
22. Saeed, N., Ridzuan, M., Maani, F.A., Alasmawi, H., Nandakumar, K., Yaqub, M.: Survrnc: Learning ordered representations for survival prediction using rank-n-contrast. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 659–669. Springer (2024)

23. Sultan, A.A., Mohamed, N., Samuel, L.T., Chughtai, M., Sodhi, N., Krebs, V.E., Stearns, K.L., Molloy, R.M., Mont, M.A.: Classification systems of hip osteonecrosis: an updated review. International orthopaedics **43**, 1089–1095 (2019)
24. Tang, W., Zhang, H., Yu, P., Kang, H., Zhang, R.: Mmmna-net for overall survival time prediction of brain tumor patients. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3805–3808. IEEE (2022)
25. Van Belle, V., Pelckmans, K., Suykens, J.A., Van Huffel, S.: Survival svm: a practical scalable algorithm. In: ESANN. pp. 89–94 (2008)
26. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. Cancer research **77**(21), e104–e107 (2017)
27. Warner, E., Lee, J., Hsu, W., Syeda-Mahmood, T., Kahn Jr, C.E., Gevaert, O., Rao, A.: Multimodal machine learning in image-based and clinical biomedicine: Survey and prospects. International Journal of Computer Vision **132**(9), 3753–3769 (2024)
28. Zhang, R., Kong, T., Wang, X., You, M.: Mask encoding: A general instance mask representation for object segmentation. Pattern Recognition **124**, 108505 (2022)
29. Zhang, R., Tian, Z., Shen, C., You, M., Yan, Y.: Mask encoding for single shot instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10226–10235 (2020)