Modeling Progression of Microwave Ablation Volumes by Neural Surrogates

Michael Holtz[1], Frankangel Servin[2], Jon S. Heiselman[2], Michael I. Miga[2], and Linwei Wang[1]

[1] Rochester Institute of Technology, Rochester NY 14623, USA
mjh9122@rit.edu
[2] Vanderbilt University, Nashville TN 37235, USA

Abstract. Microwave ablation (MWA) is a first-line treatment for early-stage hepatocellular carcinoma, but planning a probe placement that achieves an adequate ablative margin relies on biophysical simulations that take hours to run, limiting the broad search over candidate placements that planning requires. Neural surrogates can predict simulated ablation zones in a fraction of the time, but predicting the full progression of the ablation zone, not a subset of frames, is what would let such a surrogate inform decisions like ablation duration. We study three strategies for representing time: training a separate model per target time (single-frame), predicting the entire sequence at once (full-sequence), and conditioning a single model on an arbitrary query time (time-conditioned). We compare across four existing neural architectures and GRAND, a graph neural-diffusion model with an explicit temporal inductive bias, on a 218-configuration COMSOL dataset of *ex vivo* MWA. Across the full trajectory, single-frame models are most accurate at early timepoints but converge with the other strategies by mid-trajectory. Time-conditioning closes most of this early-frame gap for the weakest architectures (ED-CNN, UNETR) without materially changing the strongest ones (U-Net, Attention U-Net), while full-sequence modeling consistently biases predictions toward under-estimating final ablation volume. GRAND, the only model with explicit temporal inductive bias, underperforms the static architectures but is the only model able to accurately capture the margin of the two ablation zones.

Keywords: Microwave Ablation · Digital Twin · Neural Surrogate · Graph Neural Diffusion

1 Introduction

Liver cancer is among the deadliest cancers in the United States, with an estimated 42,340 new cases and 30,980 deaths in 2026 [17; 18]. Since resection and transplantation are available to only a limited number of patients [18], thermal ablation has become the treatment of choice for early-stage hepatocellular carcinoma (HCC) $\leq$ 3 cm [18]. Whether ablation succeeds hinges on the minimal

ablative margin (MAM): A MAM of less than 5 mm is associated with local recurrence, making a margin above 5 mm the accepted target [7; 10; 11]. Achieving this margin reliably is the central difficulty of ablation treatment.

The gap between predicted and realized ablation zones has motivated the development of patient-specific digital twins for MWA planning. Medical images are segmented into three-dimensional digital models of the liver, tumor and vasculature, over which a biophysical simulation simulates the entire ablation process [7; 15]. The result is a predicted ablation zone specific to that patient and probe placement, allowing practitioners to evaluate a plan before committing to it. This opens the door to optimization in the planning process: rather than evaluating a single probe placement, one could search thousands of configurations for the one that maximizes MAM while minimizing healthy-tissue damage and ablation time. In practice, however, the computational cost of these simulations limits their use in planning [15]. This bottleneck can be alleviated by training *neural-network surrogates* to predict the output of the simulations: by reducing the inference times from hours to milliseconds [3; 16], these surrogates make it possible to predict the ablation zones from many thousands of candidate configurations in a relatively short time frame.

Prior neural surrogates for thermal ablation predict ablation zones (*e.g.*, represented by a temperature field or ablation mask) either at a single target frame [16], or a set of predetermined time frames [3]. This is sufficient for planning at pre-chosen ablation durations, but a surgeon adjusting ablation duration intraoperatively needs the zone's progression over arbitrary time frames. Determining the appropriate stopping time precisely is of the utmost importance: stopping too early risks recurrence, while stopping too late causes unnecessary damage to healthy liver tissue. Modeling the full trajectory raises the question of how time should be represented in the model. A model could be retrained for each time point of interest (single-frame), predict the full sequence at set time points in one shot (full-sequence), or take a query time as an input and be evaluated at arbitrary points (time-conditioned). These three strategies make different assumptions about how ablation dynamics generalize over time, and it is not established which is preferable, nor whether the answer depends on the underlying architecture. In this work we explore these two questions, how to model time and which architecture to model it with, and evaluate how they interact in determining ablation zone trajectory. We study them across four static architectures adapted from RFA surrogate work, and GRAND, a graph neural diffusion model with explicit continuous time information.

2 Methodology

Signed Distance Field Representation: In this paper, we represent the ablation zones in the form of signed distance fields (SDFs). A signed distance field describes the relationship between each point on a grid and the ablation volume. At a given grid point, the signed distance field is defined as the signed Euclidean distance to the closest boundary point. Points lying inside the ablation zone are

assigned negative values, points lying directly on the boundary are assigned 0, and points outside the boundary are assigned positive. The predicted boundary can be calculated from an SDF by locating the region where predicted values flip sign.

Time Modeling Strategies: Each architecture is trained under three strategies for representing time, where the input SDF, core model architecture, and parameter budget are fixed within each comparison. In the *single frame* strategy, a separate model is trained to predict the SDF at one fixed future time. This lets a model dedicate its full parameter budget to a single timepoint, but predicting a new time of interest requires training an entirely new model. In the *full-sequence* strategy, a single model predicts the SDFs of all future time points simultaneously from one forward pass, learning the entire temporal evolution of the zone at once rather than a target time. While more practical for trajectory prediction than *single frame*, this strategy still requires points of interest to be known at train time. In the *time-conditioned* strategy, a single model predicts the SDF at an arbitrary future time supplied as an additional input, rather than a fixed set of times. Repeated inference at different query times allows one model to reconstruct the full trajectory, at arbitrary time resolution.

Model Architectures: We adapt four architectures from existing neural surrogates for RFA [3; 16] which map a tumor segmentation and probe location to a temperature field or ablation mask. Here we repurpose the same architectures to map an initial SDF to a future SDF or SDFs, including: EDCNN, a plain convolutional encoder–decoder with a compressive bottleneck [1]; U-Net, which adds skip connections that carry high resolution features past the bottleneck [8; 13]; Attention U-Net, which replaces convolution with self attention in the bottleneck and first decoder stage; and UNETR, which replaces the full convolutional encoder with a ViT [5; 6]. Their time-conditioned versions use FiLM layers to modify the outputs of each convolutional layer based on the specified target time. UNETR's time-conditioning was only applied to its convolutional decoder.

Because none of these methods has an explicit temporal-modeling component, we additionally include Graph Neural Diffusion (GRAND) which frames the problem as a continuous diffusion process over a graph, governed by a differential equation and integrated by a numerical ODE solver [2]. The diffusion is realized through an attention mechanism: at each integration step, neighboring nodes exchange messages weighted by their attention scores:

$$\frac{d\mathbf{z}_i(t)}{dt} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \, \frac{\mathbf{z}_j(t) - \mathbf{z}_i(t)}{\|\mathbf{e}_{ij}\|}, \tag{1}$$

where the graph is constructed with grid points as nodes, each one connected to its 26 nearest neighbors. $\mathbf{z}_i(t) \in \mathbb{R}^d$ is node i's latent vector at integration time t, $\mathcal{N}(i)$ is its neighborhood, $\mathbf{e}_{ij}$ is the edge vector with length $\|\mathbf{e}_{ij}\| \in \{1, \sqrt{2}, \sqrt{3}\}$ grid units, and α_{ij} is a multi-head dot-product attention weight between node i

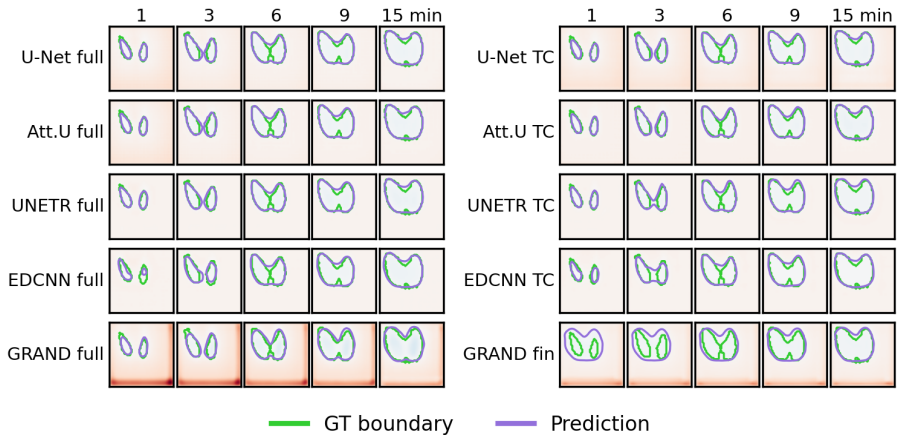

Fig. 1. Predicted ablation boundary trajectories (purple) vs. reference COMSOL boundaries (green), shown at regular intervals for each model and modeling strategy.

and node j. The dynamics are integrated in a latent space. A small convolutional encoder maps the SDF to the initial node states $\mathbf{z}_i(0)$. From these states, the diffusion process is integrated forward in time by a numerical ODE solver, and a CNN decoder maps the latent states back to a SDF. Learning a differential equation has two advantages: predictions can be made at any time in the ablation process, and the solver's step size can be adjusted at inference time, trading accuracy for computation time.

3 Experiments

Data: The dataset used is a simulation of microwave ablation in *ex vivo* liver tissue for multi-probe strategies [14]. 218 unique probe locations are generated and fed into a COMSOL (COMSOL Inc., Burlington, MA) simulation that models 15 minutes of maximum power (60 W) ablation. The simulation describes the electromagnetic waves generated by the probes via Maxwell's equations in the frequency domain, biological and electromagnetic heat transfer via Pennes' bioheat equation, and thermal tissue injury via the Arrhenius Damage Integral [14]. The simulation is sampled every 15 seconds, starting at 1 minute and ending at 15 minutes. The ablation boundary is defined by a threshold on the simulated tissue damage, giving 57 ablation boundaries per probe configuration. From these boundaries, a signed distance field (SDF) is constructed over a $10 \times 10 \times 10$ cm^3 grid centered on the ablation volume. The grid consists of evenly spaced points 2 mm apart, for a total of 132,651 spatial points. The reference sequence in Figure 1 shows the progression of the ablation over time in the COMSOL simulation.

Table 1. Accuracy averaged over the full ablation trajectory (0–15 min). TMS. = Time Modeling Strategy (single-frame (SF) vs. full-sequence (FS) vs. time-conditioned (TC)). SF rows are the single-frame models (2, 4, 8, 12, and 15 min) averaged over their target frames; FS/TC rows average every frame of the full trajectory. All entries are mean ± std across the five data splits. RVD is signed (% of reference volume; negative = under-prediction); RDD is the predicted/reference diameter ratio (1.0 = perfect) for the transverse (T) and longitudinal (L) axes; HD mm; IT is average inference time in ms. **Bold** marks the best value in each column. Underline marks the strongest of the models within each modeling strategy.

TMS.	Model	DSC ↑	IoU ↑	RVD (%)	RDD_T	RDD_L	HD ↓	IT ↓
SF	U-Net	**0.955** ± 0.002	**0.914** ± 0.003	−1.3 ± 0.6	0.972 ± 0.002	0.981 ± 0.002	**3.97** ± 0.17	**0.80** ± 0.01
SF	Att. U-Net	0.953 ± 0.001	0.910 ± 0.002	−0.6 ± 0.9	0.971 ± 0.003	0.980 ± 0.004	4.15 ± 0.17	5.18 ± 0.02
SF	UNETR	0.951 ± 0.002	0.907 ± 0.005	−0.7 ± 1.2	0.960 ± 0.004	0.976 ± 0.001	4.32 ± 0.24	53.22 ± 0.05
SF	EDCNN	0.943 ± 0.003	0.894 ± 0.006	−0.6 ± 1.0	**0.977** ± 0.004	**0.983** ± 0.005	4.35 ± 0.16	1.62 ± 0.01
FS	U-Net	0.954 ± 0.001	0.913 ± 0.002	−1.5 ± 0.5	0.963 ± 0.002	0.979 ± 0.004	4.09 ± 0.16	1.28 ± 0.08
FS	Att. U-Net	0.951 ± 0.002	0.908 ± 0.004	−0.8 ± 0.6	0.972 ± 0.003	0.981 ± 0.002	4.18 ± 0.20	5.65 ± 0.02
FS	UNETR	0.940 ± 0.019	0.899 ± 0.019	−2.5 ± 3.2	0.952 ± 0.024	0.962 ± 0.021	4.10 ± 0.20	53.45 ± 0.04
FS	EDCNN	0.914 ± 0.011	0.863 ± 0.009	−4.1 ± 1.7	0.942 ± 0.014	0.955 ± 0.011	4.62 ± 0.22	1.86 ± 0.01
TC	U-Net	0.954 ± 0.002	0.912 ± 0.003	−1.2 ± 0.8	0.968 ± 0.003	0.982 ± 0.002	4.11 ± 0.20	78.83 ± 1.21
TC	Att. U-Net	0.953 ± 0.002	0.912 ± 0.004	**−0.1** ± 1.0	0.972 ± 0.002	0.980 ± 0.004	4.06 ± 0.21	281.99 ± 0.19
TC	UNETR	0.952 ± 0.002	0.909 ± 0.004	−0.4 ± 0.9	0.957 ± 0.002	0.975 ± 0.003	4.34 ± 0.26	310.04 ± 0.12
TC	EDCNN	0.942 ± 0.003	0.892 ± 0.006	−0.7 ± 2.5	0.969 ± 0.007	0.982 ± 0.009	4.57 ± 0.22	141.63 ± 0.05
All	GRAND	0.941 ± 0.012	0.890 ± 0.021	−0.7 ± 0.8	0.963 ± 0.009	0.981 ± 0.004	4.53 ± 0.46	1038.36 ± 1.60
Final	GRAND	0.756 ± 0.010	0.649 ± 0.013	102.9 ± 6.9	1.045 ± 0.013	1.187 ± 0.014	8.56 ± 0.41	1032.44 ± 2.88

Architecture Details: GRAND uses 128-dimensional latent node features and four attention heads. The diffusion equation is integrated over the normalized interval $[0, 1]$ by the Dormand–Prince adaptive Runge–Kutta solver (`dopri5`) with relative and absolute tolerances of 10^{-3} and 10^{-4}. Gradients are propagated through the solver with the adjoint method. The full-sequence and final-frame predictions are produced from the same integration, decoded either at the final time only, or at the intermediate times corresponding to the reference frames. All optimization settings match those of the other models, except the batch size, which is set to 2 rather than 4 due to memory constraints.

Experimental Settings: All five models are trained under an identical optimization protocol on 60% train, 20% validation, 20% test splits of the 218 simulated configurations (130 training, 44 validation, 44 test). Training was repeated 5 times, with each 20% split acting as the validation once, the test set once, and training in the other three configurations. Each model minimizes a modified $\mathcal{L}_2$ loss between its predicted and reference SDFs, using Adam (learning rate 10^{-3}, weight decay 10^{-4}) with gradients clipped to a maximum norm of 1. The $\mathcal{L}_2$ loss is clamped using $\delta = 6$ to focus network capacity near the ablation boundary as in DeepSDF [12]. The learning rate is halved whenever the validation loss fails to improve for five epochs. Training runs for at most 100 epochs with early stopping after 10 epochs without validation improvement, and the checkpoint with the lowest validation loss is restored and used to calculate test set metrics. All models are trained with approximately 1.5 million parameters (1,401,187 – 1,829,169) to mitigate the effects that may come with extra capacity. GRAND's parameter count is identical under both of its supervision regimes. Other models

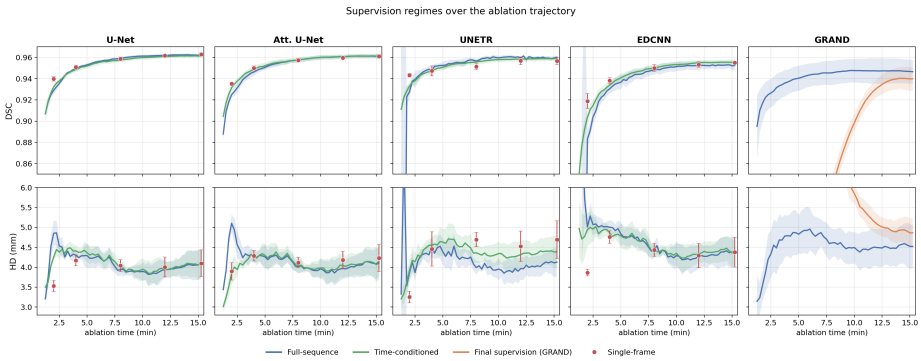

Fig. 2. DSC and HD over time for full-sequence, single-frame, and time-conditioned regimes. GRAND shows full supervision vs. final supervision. Single-frame models are trained at 2, 4, 8, 12, and 15 minutes.

vary slightly between single frame, time-conditioned, and full-sequence prediction regimes.

Metrics: We report size-, overlap-, and boundary-based metrics. The size-based relative diameter divergence (RDD, per longitudinal and transverse axis) and relative volume divergence (RVD) express the predicted–reference discrepancy as a fraction of the reference [19]. Both are scale-independent but ignore shape and location, such that a displaced or misshapen zone can still score well if its dimensions agree. The overlap-based Dice coefficient (DSC) and Jaccard index (IoU) reward agreement in size and position but not where boundary disagreement falls [4; 9]. The Hausdorff distance (HD) is the largest nearest-point distance between predicted and reference boundaries. It captures worst-case disagreement, which is especially relevant given the MAM–recurrence link [10]: a large HD marks a region where prediction and reference disagree on whether tissue is ablated, especially dangerous when it overlaps the tumor.

Prediction quality over time: DSC and HD over time are shown in Figure 2. Interestingly, all architectures and time-modeling strategies struggle most with early frames, with DSC improving as ablation time increases. As illustrated in Figure 1, the ablation volumes grow outwards from the ablation probes, merging, and forming a single ablated region. Therefore, early frames contain multiple small volumes and complex merge dynamics. The predicted boundaries struggle to time the merge accurately, with UNETR-TC and EDCNN-TC predicting a single volume at 3 minutes, while all models incorrectly predict a merged volume at the 6-minute step. Later frames contain a single contiguous volume, with boundaries largely growing away from one another rather than merging. None of the models accurately identified the small un-ablated region that persisted to the final time step. This kind of error, a false positive positioned centrally in the ablation zone, could be particularly harmful if the zone is centered on a tumor.

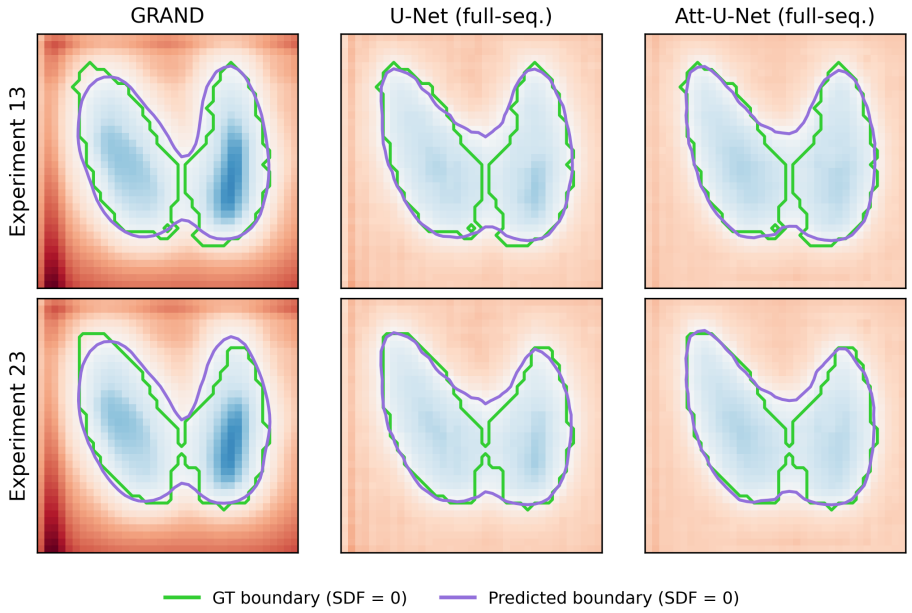

Fig. 3. Predicted and Reference boundaries on selected test configurations at the 6-minute mark.

GRAND appears qualitatively better than the other models at the 6-minute step, predicting a smaller erroneous merge region. Figure 3 shows this prediction in more detail, where GRAND captures the complex merge characteristics more accurately than the top two performing architectures.

Comparison of modeling strategies: Figure 2 shows that single-frame models outperform the other modeling strategies on the early time frames, improving DSC at the 2 minute mark. The performance gap shrinks over time: by the final time frame, DSC, IoU, and HD gaps are within a single standard deviation. Note that, in order to deploy a single frame model, the practitioner must know precisely their time of interest, and have a model trained for that frame. This is impractical as a tool for planning ablation time. Furthermore, single frame models use their full $\sim$1.5 million parameter capacity on a single frame, whereas full-sequence and time conditioned models must predict many or arbitrary frames, giving single-frame models a capacity advantage.

In comparison, full-sequence modeling seems to overall bias the models into under-predicting ablation zone's volume, as indicated by the negative RVD value. Figure 2 further shows that UNETR and EDCNN struggle with early frame prediction. Table 1 also shows much noisier metrics and large volume underestimation. This points to a failure mode explored in Figure 4, where UNETR and EDCNN collapse in the earliest frames, predicting no volume at all. The time-

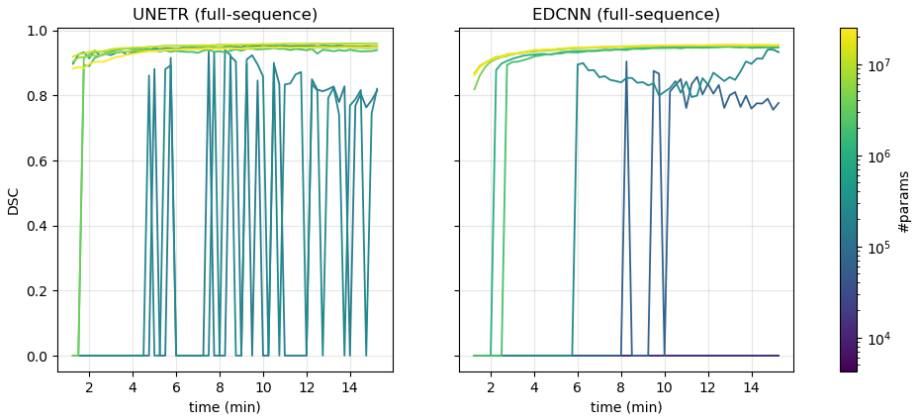

Fig. 4. Dice Similarity Coefficient over time for UNETR and EDCNN architectures as number of trainable parameters vary.

conditioned variants of these architectures do not suffer from the same issues. Figure 4 shows the relationship between number of parameters and the DSC over time. While large models show the expected relationship between DSC and time, smaller models either do not predict any volume for early time steps, or predict volumes only for certain early times. This leads to the conclusion that time-conditioned variants have a large impact on small or weak models, but stronger models do not benefit substantively over the full-sequence prediction.

Comparison across architectures: U-Net and Att. U-Net form the strongest architecture cluster across the full trajectory (Table 1), regardless of modeling strategy, with U-Net edging ahead on overlap and boundary metrics. EDCNN is consistently the worst of the static models and struggles most under full supervision. Inference cost varies widely (Table 1): U-Net is fastest, time-conditioning is far slower due to repeated forward passes, and GRAND slower still from ODE solving, though all are dwarfed by the hours a simulation takes.

Additional ablation on GRAND: We conduct additional ablation studies to supervise GRAND solely on the final timestep. Table 1 and Figure 1 show that supervising the final timestep alone results in both worse predictions on the final step, and far worse predictions on the rest of the trajectory. Without intermediate reference frames, GRAND's full trajectory DSC collapses, and RVD balloons to 102.9%, far worse than any other model. The trajectory plot shows that GRAND's final-frame model predicts a similar volume for all time steps, implying that GRAND's continuous-time formulation and diffusion mechanism are not sufficient to recover the full trajectory by themselves.

4 Discussion and Conclusion

Several factors limit these results and point to extensions. The dataset is *ex vivo*, omitting the tumor, vasculature, and fat heterogeneity that shape patient ablation zones. Conditioning on segmented patient-specific geometry alongside probe placement would let the models evaluate plans from imaging rather than forecast a simulation in progress. GRAND's graph formulation is a natural fit for such irregular anatomy. The models predict from an SDF sampled one minute in rather than from initial conditions; predicting directly from probe locations would remove the need to simulate that first minute. Finally, 2 mm sampling limits boundary localization and causes saturation on boundary metrics, motivating the need for finer grids and direct prediction of tissue-damage fields. Jointly modeling multiple simulation variables is a further extension.

We compared single-frame, time-conditioned, and full-sequence time modeling across four static neural surrogates and GRAND for predicting microwave ablation zone evolution from a 218-configuration COMSOL dataset. Architecture mattered more than modeling strategy at the final frame, with U-Net and Att. U-Net strongest throughout. The weaker EDCNN and UNETR were regime-sensitive, showing an early frame volume collapse under full-sequence modeling that time-conditioning largely resolved. Full sequence modeling consistently biased final-volume predictions downward, and single frame models led early but converged with others by later time points. GRAND, the only architecture with an explicit temporal inductive bias, underperformed the static architectures but excelled in complex merge areas.

References

[1] Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (Dec 2017). https://doi.org/10.1109/TPAMI.2016.2644615

[2] Chamberlain, B.P., Rowbottom, J., Gorinova, M., Webb, S., Rossi, E., Bronstein, M.M.: GRAND: Graph Neural Diffusion (Sep 2021). https://doi.org/10.48550/arXiv.2106.10934

[3] Cho, S., Seo, M., Shin, M., Yoon, K.: Towards Digital Twin of RF Ablation: Real-Time Prediction of Time-Dependent Thermal Effects Using Transformer. In: Li, L., Jirsa, V., Feng, J., Deng, J., Dede', L., An, S., Lyu, Y., Liu, X. (eds.) Digital Twin for Healthcare, vol. 16193, pp. 69–78. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-07694-6_7

[4] Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology **26**(3), 297–302 (Jul 1945). https://doi.org/10.2307/1932409

[5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (Jun 2021). https://doi.org/10.48550/arXiv.2010.11929

[6] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H., Xu, D.: UNETR: Transformers for 3D Medical Image Segmentation (Oct 2021). https://doi.org/10.48550/arXiv.2103.10504

[7] Heshmat, A., O'Connor, C.S., Albuquerque Marques Silva, J., Paolucci, I., Jones, A.K., Odisio, B.C., Brock, K.K.: Using Patient-Specific 3D Modeling and Simulations to Optimize Microwave Ablation Therapy for Liver Cancer. Cancers **16**(11), 2095 (May 2024). https://doi.org/10.3390/cancers16112095

[8] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Kohler, G., Norajitra, T., Maier-Hein, K.H.: nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation

[9] Jaccard, P.: THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.[1]. New Phytologist **11**(2), 37–50 (Feb 1912). https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

[10] Laimer, G., Schullian, P., Jaschke, N., Putzer, D., Eberle, G., Alzaga, A., Odisio, B., Bale, R.: Minimal ablative margin (MAM) assessment with image fusion: An independent predictor for local tumor progression in hepatocellular carcinoma after stereotactic radiofrequency ablation. European Radiology **30**(5), 2463–2472 (May 2020). https://doi.org/10.1007/s00330-019-06609-7

[11] Lin, Y.M., Paolucci, I., O'Connor, C.S., Anderson, B.M., Rigaud, B., Fellman, B.M., Jones, K.A., Brock, K.K., Odisio, B.C.: Ablative Margins of Colorectal Liver Metastases Using Deformable CT Image Registration and Autosegmentation. Radiology **307**(2), e221373 (Apr 2023). https://doi.org/10.1148/radiol.221373

[12] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation (Jan 2019). https://doi.org/10.48550/arXiv.1901.05103

[13] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (May 2015). https://doi.org/10.48550/arXiv.1505.04597

[14] Servin, F., Collins, J., Heiselman, J.S., Planz, V.B., Kolouri, S., Brown, D.B., Miga, M.I.: Digital twin modeling and machine learning frameworks for forecasting multiple microwave ablation volumes. In: Rettmann, M.E., Siewerdsen, J.H. (eds.) Medical Imaging 2025: Image-Guided Procedures, Robotic Interventions, and Modeling. p. 71. SPIE, San Diego, United States (Apr 2025). https://doi.org/10.1117/12.3047302

[15] Servin, F., Collins, J.A., Heiselman, J.S., Frederick-Dyer, K.C., Planz, V.B., Geevarghese, S.K., Brown, D.B., Jarnagin, W.R., Miga, M.I.: Simulation of Image-Guided Microwave Ablation Therapy Using a Digital Twin Computational Model. IEEE Open Journal of Engineering in Medicine and Biology **5**, 107–124 (2024). https://doi.org/10.1109/OJEMB.2023.3345733

[16] Shin, M., Seo, M., Cho, S., Park, J., Kwon, J.H., Lee, D., Yoon, K.: PhysRFANet: Physics-guided neural network for real-time prediction of thermal effect during radiofrequency ablation treatment. Engineering Applications of Artificial Intelligence **138**, 109349 (Dec 2024). https://doi.org/10.1016/j.engappai.2024.109349

[17] Siegel, R.L., Kratzer, T.B., Wagle, N.S., Sung, H., Jemal, A.: Cancer statistics, 2026. CA: A Cancer Journal for Clinicians **76**(1), e70043 (Jan 2026). https://doi.org/10.3322/caac.70043

[18] Singal, A.G., Llovet, J.M., Yarchoan, M., Mehta, N., Heimbach, J.K., Dawson, L.A., Jou, J.H., Kulik, L.M., Agopian, V.G., Marrero, J.A., Mendiratta-Lala, M., Brown, D.B., Rilling, W.S., Goyal, L., Wei, A.C., Taddei, T.H.: AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology **78**(6), 1922–1965 (Dec 2023). https://doi.org/10.1097/HEP.0000000000000466

[19] Van Erp, G.C.M., Hendriks, P., Broersen, A., Verhagen, C.A.M., Gholamiankhah, F., Dijkstra, J., Burgmans, M.C.: Computational Modeling of Thermal Ablation Zones in the Liver: A Systematic Review. Cancers **15**(23), 5684 (Dec 2023). https://doi.org/10.3390/cancers15235684