Received: 21 March 2022 | Revised: 27 July 2022 | Accepted: 8 November 2022

DOI: 10.1002/mp.16181

RESEARCH ARTICLE

MEDICAL PHYSICS

Towards optimal deep fusion of imaging and clinical data via a model-based description of fusion quality

Yuqi Wang[1] | **Xiang Li**[1] | **Meghana Konanur**[2] | **Brandon Konkel**[2] | **Elisabeth Seyferth**[2] | **Nathan Brajer**[2] | **Jian-Guo Liu**[3,4] | **Mustafa R. Bashir**[2,5] | **Kyle J. Lafata**[1,2,6]

[1]Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, USA

[2]Department of Radiology, Duke University, Durham, North Carolina, USA

[3]Department of Mathematics, Duke University, Durham, North Carolina, USA

[4]Department of Physics, Duke University, Durham, North Carolina, USA

[5]Department of Medicine, Gastroenterology, Duke University, Durham, North Carolina, USA

[6]Department of Radiation Oncology, Duke University, Durham, North Carolina, USA

Correspondence
Kyle J. Lafata, Radiology, Radiation Oncology, and Electrical & Computer Engineering, Duke University, Durham, NC, USA.
Email: kyle.lafata@duke.edu

Abstract

Background: Due to intrinsic differences in data formatting, data structure, and underlying semantic information, the integration of imaging data with clinical data can be non-trivial. Optimal integration requires robust data fusion, that is, the process of integrating multiple data sources to produce more useful information than captured by individual data sources. Here, we introduce the concept of *fusion quality* for deep learning problems involving imaging and clinical data. We first provide a general theoretical framework and numerical validation of our technique. To demonstrate real-world applicability, we then apply our technique to optimize the fusion of CT imaging and hepatic blood markers to estimate portal venous hypertension, which is linked to prognosis in patients with cirrhosis of the liver.

Purpose: To develop a measurement method of optimal data fusion quality deep learning problems utilizing both imaging data and clinical data.

Methods: Our approach is based on modeling the fully connected layer (FCL) of a convolutional neural network (CNN) as a potential function, whose distribution takes the form of the classical Gibbs measure. The features of the FCL are then modeled as random variables governed by state functions, which are interpreted as the different data sources to be fused. The probability density of each source, relative to the probability density of the FCL, represents a quantitative measure of source-bias. To minimize this source-bias and optimize CNN performance, we implement a vector-growing encoding scheme called positional encoding, where low-dimensional clinical data are transcribed into a rich feature space that complements high-dimensional imaging features. We first provide a numerical validation of our approach based on simulated Gaussian processes. We then applied our approach to patient data, where we optimized the fusion of CT images with blood markers to predict portal venous hypertension in patients with cirrhosis of the liver. This patient study was based on a modified ResNet-152 model that incorporates both images and blood markers as input. These two data sources were processed in parallel, fused into a single FCL, and optimized based on our fusion quality framework.

Results: Numerical validation of our approach confirmed that the probability density function of a fused feature space converges to a source-specific probability density function when source data are improperly fused. Our numerical results demonstrate that this phenomenon can be quantified as a measure of fusion quality. On patient data, the fused model consisting of both imaging data and positionally encoded blood markers at the theoretically optimal fusion quality metric achieved an AUC of 0.74 and an accuracy of 0.71. This model was statistically better than the imaging-only model (AUC = 0.60; accuracy = 0.62),

the blood marker-only model (AUC = 0.58; accuracy = 0.60), and a variety of purposely sub-optimized fusion models (AUC = 0.61–0.70; accuracy = 0.58–0.69).

Conclusions: We introduced the concept of data fusion quality for multi-source deep learning problems involving both imaging and clinical data. We provided a theoretical framework, numerical validation, and real-world application in abdominal radiology. Our data suggests that CT imaging and hepatic blood markers provide complementary diagnostic information when appropriately fused.

KEYWORDS
data fusion, deep learning, imaging, radiomics

1 | INTRODUCTION

In medicine, imaging data (e.g., CT, MRI) and non-imaging data (e.g., demographics, lab results) provide complementary diagnostic utility. Combining pixel data with other clinically relevant information may lead to improved characterization of a disease. However, their computational integration is non-trivial due to intrinsic differences in data formatting, data structure, and underlying semantic information. Optimal integration therefore requires robust data fusion, that is, the process of integrating multiple data sources to produce more consistent, accurate, and useful information than captured by individual data sources. Data fusion techniques aim to improve data-driven inference and downstream modeling by assessing the association, correlation, and combination of data from multiple sources or sensors.[1,2]

The fusion of data can be categorized as *early* fusion, *joint* fusion, or *late* fusion, respectively, depending on whether the data are fused prior to, during, or after model training.[3] These techniques have been widely used in various disciplines (e.g., transportation, robotics) and are becoming increasingly relevant to medical imaging applications of deep learning.[4–13] For example, Kharazmi et al. used combined CNN extracted features of dermoscopic images and then concatenated them with patient data and genetic data as the input of a basal cell carcinoma detection model.[4] Similarly, Li et al. trained a LSTM autoencoder using cognitive assessments and patient demographics to build a compact representation, then combined it with MR images to train a prediction model for Alzheimer's disease.[5] Likewise, Yala et al. demonstrated that the fusion of a mammogram-based ResNet model with a risk-factor-based logistic regression model improved the performance of breast cancer prediction relative to the individual models.[6]

Although many of these studies report increased model performance, their data fusion is typically a heuristic process, whereby data from multiple sources are concatenated without any preprocessing or measurement of *fusion quality* other than downstream model performance. This heuristic approach lacks interpretation of the fused feature space and may cause sub-optimal model performance. In fact, data preprocessing can significantly influence the performance of a fused model due to differences in source data characteristics.[14] This is particularly relevant in medical imaging problems, where low-dimensional clinical features can often ϵ be overpowered by high-dimensional imaging features, resulting in non-equitable contributions from individual data sources to the model.

Here, we propose a new approach to quantify the quality of data fusion in multi-source deep learning problems. Briefly, our technique is based on modelling the fully connected layer (FCL) of a deep neural network as a potential function whose probability distribution takes the form of the classical Gibbs measure. The features of the FCL are modeled as random variables governed by state functions, which are interpreted as the different data sources to be fused. The contribution of each source, relative to the probability density of the FCL, represents a quantitative measure of source-bias. To minimize this source-bias and optimize data fusion quality, we implement a vector-growing encoding scheme, known as positional encoding,[15] where low-dimensional clinical features are transcribed into a rich feature space to complement high-dimensional imaging features. To demonstrate the feasibility of our technique, this paper provides: (a) a theoretical framework, (b) a numerical validation, and (c) a clinically relevant application in abdominal radiology, where CT imaging and hepatic blood markers are fused to predict portal venous hypertension in patients with cirrhosis of the liver.

2 | METHODS

2.1 | Feature density estimation and data fusion theory

2.1.1 | Probability density estimation of fused feature vectors

To characterize the fusion of features from different data sources, we model features as random variables

and data sources as state functions. We assume that a given feature vector $\boldsymbol{f} \in \mathbb{R}^d$ is sampled from a more general distribution that obeys hypotheses of a canonical ensemble. The probability distribution of $\boldsymbol{f}$ takes the form of a classical Gibbs measure,

$$\varphi \sim Z^{-1} e^{-\beta E} \tag{1}$$

where, E is the total energy of the system, β is a scaling constant, and Z is the appropriate partition function that encodes how the probabilities are partitioned among the different states.[16] Accordingly, we are motivated to interpret the different data sources of $\boldsymbol{f}$ as the state functions of E, such that states with lower energy density will always have a higher probability of being occupied.

First, we define high-dimensional imaging features as the points, $\boldsymbol{x} = \{x_i\}_{i=1}^{d_1} \in \mathbb{R}^{d_1}$, and low-dimensional clinical features as the points, $\boldsymbol{y} = \{y_j\}_{j=1}^{d_2} \in \mathbb{R}^{d_2}$. Here, $\boldsymbol{x}$ and $\boldsymbol{y}$ are each a subset of $\boldsymbol{f}$ such that $d = d_1 + d_2$ and $d_2 \ll d_1$, thus reflecting the real-world dimensional disparity between imaging features and clinical features. Next, we use a kernel density estimation technique to approximate the probability density functions of $\boldsymbol{x}$ and $\boldsymbol{y}$. The imaging features are first mapped from their original Euclidean space into a Hilbert space of square integrable functions,

$$\varphi_i(\boldsymbol{x}) = \frac{1}{d_1} \sum_{i=1}^{d_1} e^{-\frac{1}{2\sigma^2}[\boldsymbol{x}-x_i]^2} \tag{2}$$

where, φ_i is the marginal contribution of imaging parameterized by a radial basis function kernel *full-width-half-max* of σ. Similarly, the marginal contribution of the clinical features is,

$$\varphi_j(\boldsymbol{y}) = \frac{1}{d_2} \sum_{j=1}^{d_2} e^{-\frac{1}{2\sigma^2}[\boldsymbol{y}-y_j]^2}. \tag{3}$$

Based on Equations (2) and (3), the probability density function of $\boldsymbol{f}$ can be written as,

$$\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \alpha \sum_{i=1}^{d_1} e^{-\frac{1}{2\sigma^2}[\boldsymbol{x}-x_i]^2} + \beta \sum_{j=1}^{d_2} e^{-\frac{1}{2\sigma^2}[\boldsymbol{y}-y_j]^2} \tag{4}$$

where, $\alpha = d_1^{-1}$ and $\beta = d_2^{-1}$ are weighting coefficients of the relative contribution of φ_i and φ_j, respectively, on $\boldsymbol{f}$.

In general, when a mechanical system is in equilibrium with a heatbath, the system will exchange energy with the heatbath, such that the microstates of the system will differ in total energy. Therefore, when Equation (4) is in equilibrium, the distribution of data points $\boldsymbol{x}$ and $\boldsymbol{y}$ only depends on the energy difference between the two states,[17,18]

$$\Delta E = \sum_i \varphi_i^2(\boldsymbol{x}) - \sum_j \varphi_j^2(\boldsymbol{y}). \tag{5}$$

According to Equation (5), different state functions have equal probabilities of occurring when ΔE is minimized. However, when the difference in energy between states is high, $\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y})$ instead converges to a marginal contribution parameterized by the dimensions d_1 and d_2, that is,

$$d_2 \ll d_1 \rightarrow \varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}) \approx \varphi_i(\boldsymbol{x}) \tag{6}$$

and

$$d_2 \gg d_1 \rightarrow \varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}) \approx \varphi_j(\boldsymbol{y}). \tag{7}$$

Our motivating hypothesis is that fused features from different data sources can only provide complementary information if their combined probability density function is not governed by the marginal contribution of individual data sources (i.e., Equation (5) is minimized). This a priori quantification of data fusion quality enables optimization of downstream multi-source deep learning.

2.1.2 | Fusion via positional encoding

To reduce the energy of $\boldsymbol{x}$ and ensure that neither Equations (6) nor (7) are satisfied, we utilize a positional encoding procedure to extend the dimension of $\boldsymbol{y}$ from its native d_2 into an embedding space of dimension d_{2^*}. Positional encoding is described as,

$$PE(\boldsymbol{y},\ 2k|d_{2^*}) = sin\left(\frac{\boldsymbol{y}}{10000^{\frac{2k}{d_{2^*}}}}\right) \tag{8}$$

and

$$PE(\boldsymbol{y},\ 2k+1|d_{2^*}) = cos\left(\frac{\boldsymbol{y}}{10000^{\frac{2k}{d_{2^*}}}}\right) \tag{9}$$

where, $\boldsymbol{y} = \{y_j\}_{j=1}^{d_2} \epsilon \mathbb{R}^{d_2}$ are the input positions of a set of points, d_{2^*} is the dimension of the encoding space, k is the index of the encoding space bound on the interval $[0,\ d_{2^*}/2)$, and $\boldsymbol{y}^* = PE(\boldsymbol{y}) \in \mathbb{R}^{d_{2^*}}$ is the encoded feature vector of $\boldsymbol{y}$. Positional Encoding was first proposed in the paper, *Attention Is All You Need*[15] and has been applied to various word embedding problems[19,20] and clinical data vector growing operations.[21] In general, Equations (8) and (9) provide a unique and deterministic

encoding for each value and ensures the distance between any two values is consistent.

Following positional encoding, the points, $\{y_j\}_{j=1}^{d_2} \epsilon \mathbb{R}^{d_2}$, are converted from scalars to vectors, which reduces the influence of the length imbalance relative to the points $\{x_i\}_{i=1}^{d_1} \epsilon \mathbb{R}^{d_1}$. We can therefore re-cast Equation (4) in terms of $\boldsymbol{y}^*$ and d_{2^*},

$$\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}^*) = \alpha \sum_{i=1}^{d_1} e^{-\frac{1}{2\sigma^2}[\boldsymbol{x}-x_i]^2} + \beta^* \sum_{j=1}^{d_2} e^{-\frac{1}{2\sigma^2}[\boldsymbol{y}^*-PE(y_j|d_{2^*})]^2} \quad (10)$$

where, $PE(y_j|d_{2^*})$ is the response of the positional encoding procedure on y_j into a d_{2^*}-dimensional embedding space according to Equations (8) and (9), and β^* is a re-weighting of β in d_{2^*}-space. Since $\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}^*)$ is parameterized by both d_1 and d_{2^*}, we define a scaling coefficient,

$$\gamma = \frac{\alpha}{\beta^*} = \frac{d_{2^*}}{d_1}, \quad (11)$$

which represents data fusion quality. The metric γ can be used to easily characterize the response of different positional encoding implementations and its relative effect on ΔE. That is, $\Delta E \approx 0$ when $\gamma \approx 1$.

2.1.3 | Numerical validation

To characterize Equation (10) and verify Equations (6) and (7), we performed a numerical analysis. The following computational experiments were performed to verify our intuition and to better understand the parameters driving equitable data fusion. We modeled the points, $\boldsymbol{x} = \{x_i\}_{i=1}^{d_1} \in \mathbb{R}^{d_1}$, as a Gaussian process with $d_1 = 1000$. The points, $\boldsymbol{y} = \{y_j\}_{j=1}^{d_2} \in \mathbb{R}^{d_2}$, were then modeled by sparsely sampling from the Gaussian distribution with 1000 stochastic iterations and $d_2 = 3$. At each iteration, positional encoding was applied to $\boldsymbol{y}$ according to Equations (8) and (9) for a given d_{2^*} value. The result was averaged across all iterations to generate a d_{2^*}-dimensional embedding space, $\boldsymbol{y}^*$. The vectors $\boldsymbol{x} \in \mathbb{R}^{d_1}$ and $\boldsymbol{y}^* \in \mathbb{R}^{d_{2^*}}$ were concatenated to generate a fused feature space, $\boldsymbol{f} \in \mathbb{R}^{1000+d_{2^*}}$.

To study the effect of d_{2^*} on $\boldsymbol{f}$, we chose monotonically increasing values of $d_{2^*} = \{10, 100, 500, 1000, 10000, 100000\}$. The probability density functions of $\boldsymbol{x}$ and $\boldsymbol{y}^*$ were calculated according to Equations (2) and (3), respectively, to generate $\varphi_i(\boldsymbol{x})$ and $\varphi_j(\boldsymbol{y}^*)$. For each d_{2^*} value, the joint probability, $\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}^*)$, was calculated according to Equation 10 and characterized via γ values according to Equation (11). While $\varphi_i(\boldsymbol{x})$ is invariant to changes in d_{2^*}, the net-effect on $\varphi_{ij}(\boldsymbol{x}, \boldsymbol{y}^*)$ depends on the magnitude of d_{2^*} relative to d_1. To quantify this trend, the energy difference between states, ΔE, was calculated according to Equation (5) and compared across the d_{2^*} values to numerically verify that ΔE is minimized when $\gamma \approx 1$.

2.2 | Deep fusion of CT imaging and hepatic blood markers to estimate portal venous hypertension

In this section, we apply the theory presented in Section 2.1 to a patient dataset to fuse CT images with hepatic blood markers and predict portal venous hypertension in patients with cirrhosis of the liver. Uncontrolled portal hypertension leads to numerous complications, such as gastroesophageal variceal bleeding, ascites, hepatorenal syndrome, and hepatic encephalopathy.[22] Although the extent of portal hypertension can be directly measured via the hepatic venous pressure gradient, this requires an invasive procedure and serial measurements are often not feasible. As a non-invasive alternative, both abdominal imaging and hepatic blood markers are being investigated as potential biomarkers of the hepatic venous pressure gradient.[23] These data may provide complementary knowledge and new insight, leading to improved characterization of portal hypertension in cirrhotic patients. However, for reasons stated above, integration of these data is challenging.

2.2.1 | Multi-source deep learning architecture design

We designed a deep learning pipeline (Figure 1) to predict portal venous hypertension based on the fusion of two complementary data sources: (1) CT imaging and (2) hepatic blood markers. The approach is based on a modified ResNet-152 model that incorporates both images and blood markers as input. These two data sources are processed in parallel and then fused into a single fully connected layer based on the mathematical framework described in Section 2.1. Imaging features, $\boldsymbol{x} = \{x_i\}_{i=1}^{d_1} \in \mathbb{R}^{d_1}$, are derived from the ResNet-152 model and are of dimension $d_1 = 2048$. Specifically, we chose to characterize the surface of the left hepatic lobe on CT imaging, because previous research has demonstrated an association between surface nodularity and portal pressure.[24,25] Here, the left hepatic lobe serves as a *prior* in the deep learning framework. Blood markers, $\boldsymbol{y} = \{y_j\}_{j=1}^{d_2} \in \mathbb{R}^{d_2}$, included *APRI*, *Albumin*, and *Platelet Count*, which are all known to be associated with the pathogenesis of portal hypertension. Blood markers are therefore of dimension $d_2 = 3$.

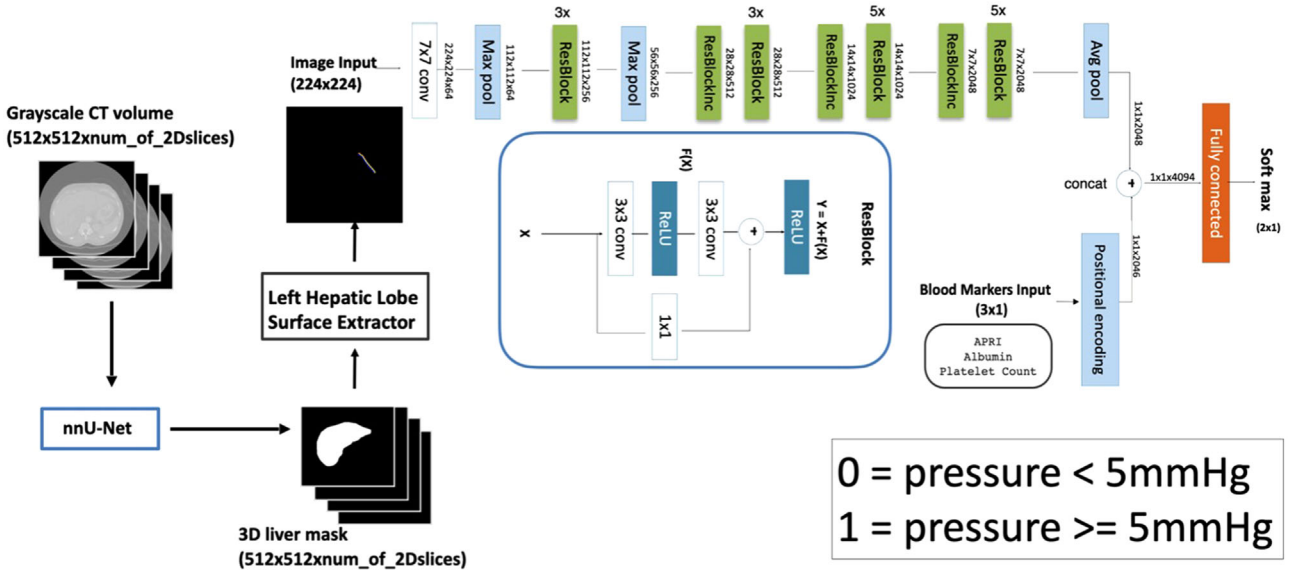

FIGURE 1 Multi-source deep learning pipeline to predict portal venous hypertension based on the fusion of two complementary data sources: (1) CT imaging and (2) hepatic blood markers.

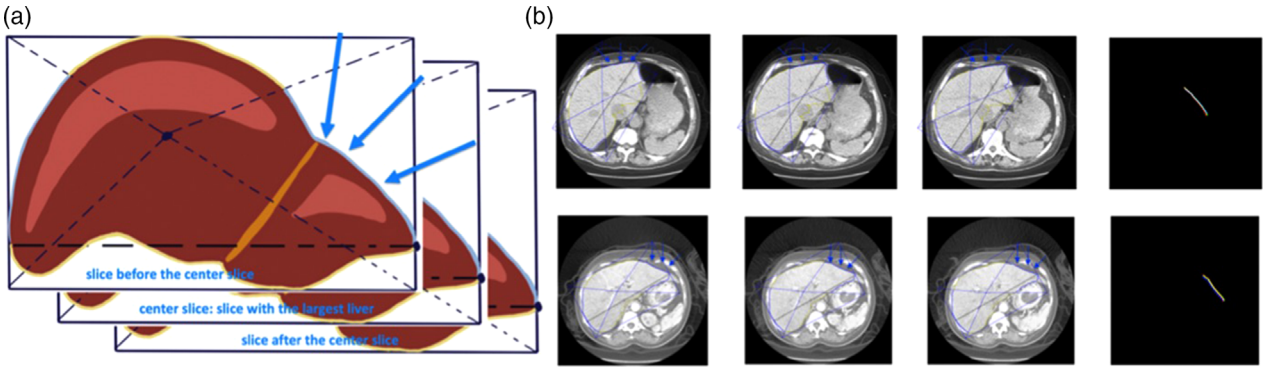

FIGURE 2 Automatic detection of the left hepatic lobe surface on CT. (a) A toy example illustrating surface characterization via liver segmentation, center-of-mass detection, and tip detection. (b) Illustrating example of the method applied to two representative abdominal CT images.

The approach is as follows. The 3D liver volume is first segmented on CT using a previously validated self-adapting nnU-Net[26,27] model. The liver tip and center-of-mass are then automatically detected as anatomic landmarks by finding the longest axial distance across the liver as illustrated by the blue dotted line on Figure 2a. The surface of the left hepatic lobe is then detected by a rotationally invariant bounding box technique to partition the surface relative to its tip and center-of-mass. The blue arrows on Figure 2a illustrate the surface of the left hepatic lobe. The 2D surface in the superior-inferior direction is then captured by combining adjacent slices into a three-channel image and rotating it to approximate a rotationally invariant system across different patients. As an illustrating example, Figure 2b demonstrates this surface detection in two different patients. Acting as a prior, this surface is passed to a pre-trained ResNet-152 model[28,29] to encode surface nodularity as the set of deep imaging features, $\boldsymbol{x} = \{x_i\}_{i=1}^{d_1} \in \mathbb{R}^{d_1}$, where $d_1 = 2048$.

In parallel to the imaging branch, positional encoding is applied to the blood marker data based on Equations (8) and (9). The rationale here is to transcribe the raw blood marker features, $\boldsymbol{y} \in \mathbb{R}^{d_2}$, into a rich feature space, $\boldsymbol{y}^* \in \mathbb{R}^{d_{2^*}}$, that better complements the fully connected layer of the ResNet-152 imaging features, $\boldsymbol{x} \in \mathbb{R}^{d_1}$. Following positional encoding of $\boldsymbol{y} \rightarrow \boldsymbol{y}^*$, the vectors $\boldsymbol{x}$ and $\boldsymbol{y}^*$ are concatenated into a single fully connected layer, $\boldsymbol{f} \in \mathbb{R}^{d_1+d_{2^*}}$, that is mapped to a diagnosis of portal hypertension, $C \in \{0,\ 1\}$, via a soft max operation.

As an illustrating example, the heatmap on Figure 3a demonstrates positionally encoded blood markers. Positional encoding provides a unique and deterministic

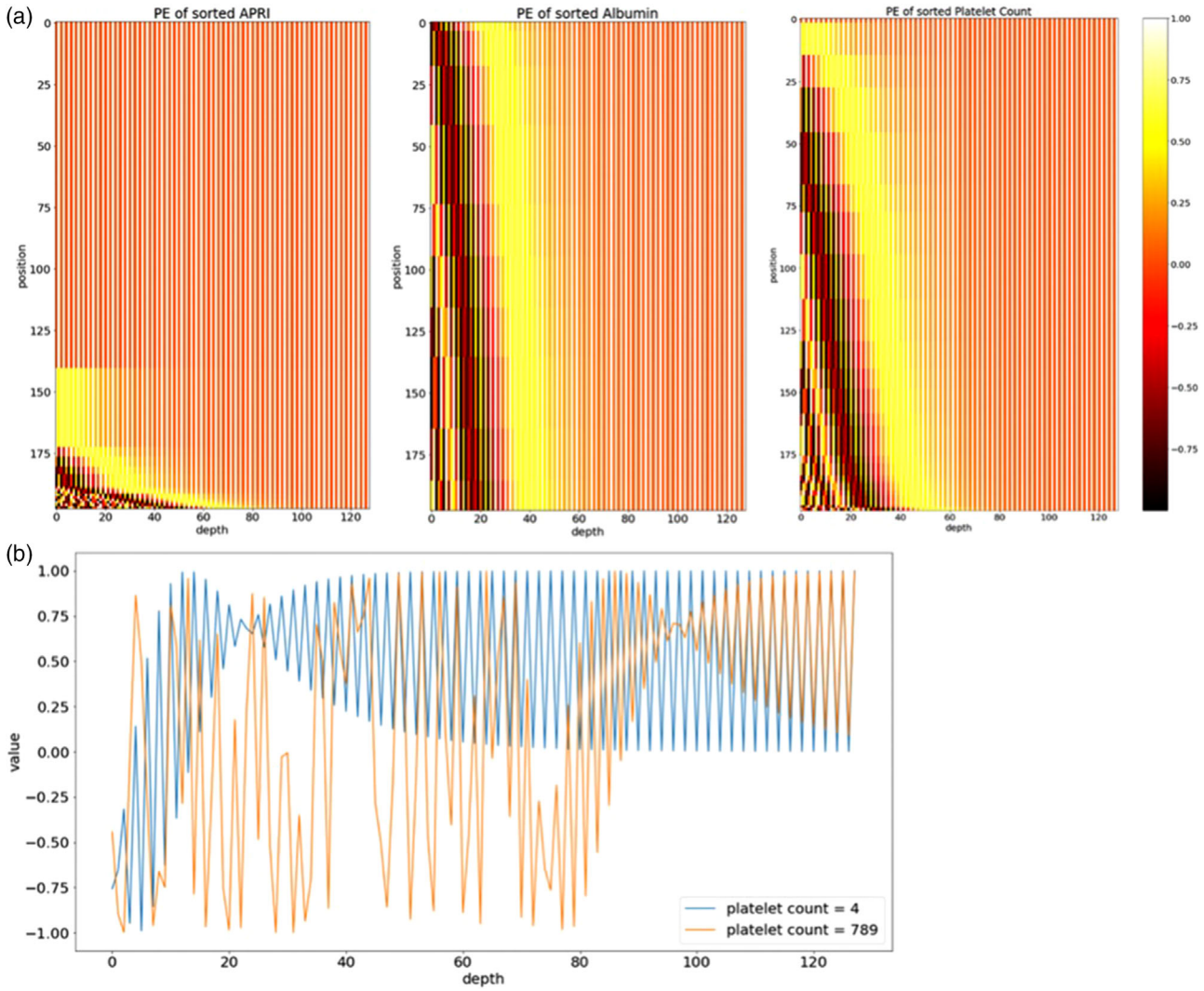

FIGURE 3 An illustrating example of positionally encoded blood markers. (a) The heatmaps demonstrate positionally encoding of APRI (left), Albumin (middle), and Platelet Count (right), where depth represents each element in the vector. (b) The positional encoding of two different platelet count values (4 K/μL vs. 789 K/μL) from two different patients is plotted at a dimension of $d_{2*} = 128$. The blue line represents a platelet count of 4 K/μL, while the orange line represents a platelet count of 789 K/μL.

encoding for each value and ensures that the distance between any two values is consistent. The value of the raw blood makers is thus represented by the position of frequency change. As shown on Figure 3b, the positional encoding of two different platelet count values (4 K/μL vs. 789 K/μL) of two different patients is plotted at a dimension of $d_{2*} = 128$. The blue line represents a platelet count of 4 K/μL, while the orange line represents a platelet count of 789 k/μL. Importantly, the encoded vector contains the positional information of the raw data, which can be harnessed to differentiate the two values in higher-dimensional feature space.

2.2.2 | Model training

To train the model described in Section 2.2.1, we retrospectively identified 198 patients at our institution with available CT imaging, blood marker laboratory results, and invasive hepatic venous pressure gradient (HVPG) measurements within 2 weeks of both image acquisition and blood draw. We excluded patients who had had prior TIPS (trans jugular intrahepatic portosystemic shunt) and prior liver transplantation. Portal venous hypertension was based on its clinical definition of HVPG $\geq$ 5 mmHg, which was used to define the class label, $C \in \{0, 1\}$. In total, 119 patients had portal venous hypertension (i.e., $C = 1$), and 79 did not (i.e., $C = 0$). The data was partitioned into training (80%), validation (10%), and testing (10%) sets using a stratified Monte Carlo sampling technique to maintain an equal class ratio for each partition. Training consisted of a batch size of eight cases, stochastic gradient descent with 0.9 momentum, and binary cross entropy loss. The initial learning rate was 0.001 with a step learning rate decrease procedure. To prevent overfitting and boost

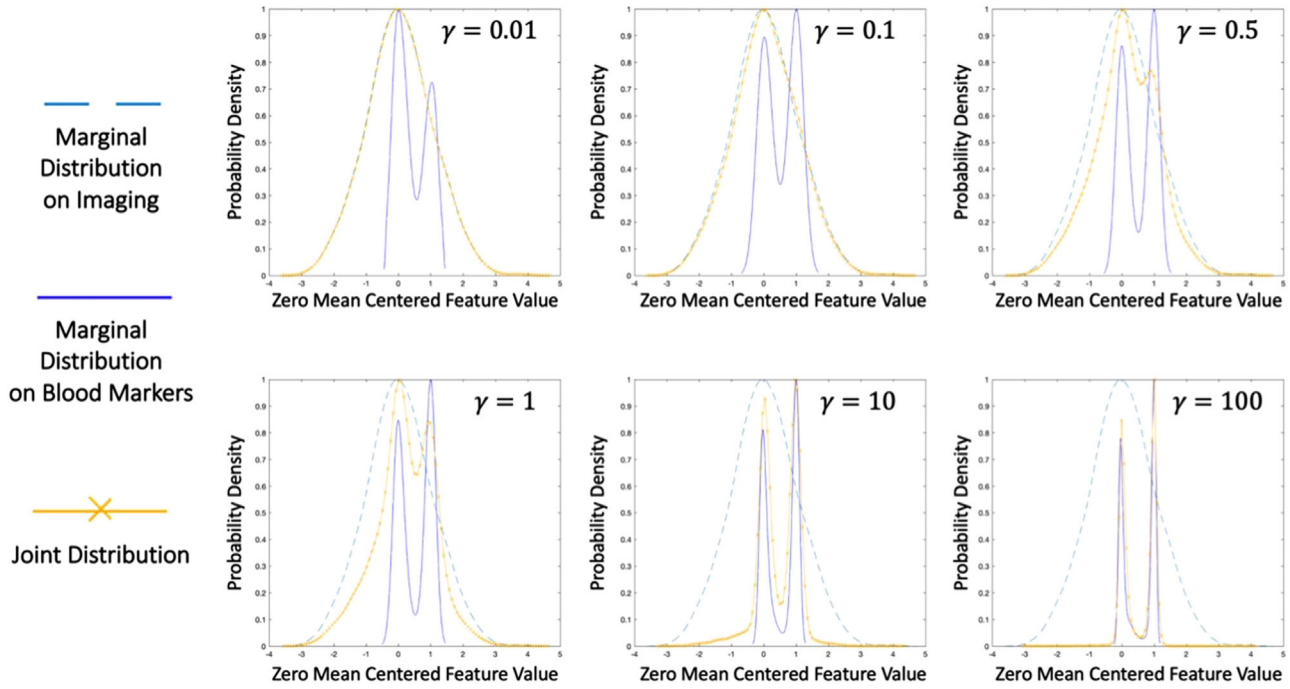

FIGURE 4 Experimental results and numerical validation of the proposed fusion measurement technique. Marginal distributions of a Gaussian process and sparsely sampled points, to simulate imaging features and blood markers, respectively, are shown relative to their joint distribution. As γ is monotonically increased, the shape of the joint distribution transitions from the Gaussian process to the sparsely sampled points. At the extremes (i.e., $\gamma = 0.01$; $\gamma = 100$) the joint distribution converges to a given sou distribution, indicating poor data fusion. When $\gamma = 1$, the joint distribution does not approximate either marginal distribution, i.e., the energy between the states is minimized. This indicates optimal data fusion, where each source domain has an equal opportunity to contribute to the model.

model generalization, we implemented label smoothing, early stopping, and horizontal flip data augmentation. The maximum iteration was set at 80 epochs.

2.2.3 | Evaluation and performance metrics

To evaluate the effect imaging plus blood maker data fusion on downstream model performance, we trained several models in parallel and compared their relative performance. These models included: (i) a CT imaging-only model; (ii) a blood marker-only model *without* positional encoding; (iii) a blood marker-only model *with* positional encoding; and (iv) several d_{2^*} specific fused imaging + blood marker models, parameterized based on the methodology proposed in Section 2.1. Model performance was based on test set receiver operating characteristic (ROC) curve analysis.

In addition, we also quantified fusion quality based on feature density estimation as a function of $d_{2^*} = \{3, 192, 384, 768, 2046, 3073, 6144\}$. This is analogous to the numerical analysis performed in Section 2.1.3. Each d_{2^*}-dependent blood marker embedding was fused with the 2048 ResNet-152 imaging features. The resulting feature vector, $\boldsymbol{f} \in \mathbb{R}^{2048+d_{2^*}}$, was used to quantify fusion quality based on the relative similarity between the marginal contributions and the probability density function of the fused FCL as previously described in Section 2.1.

3 | RESULTS

3.1 | Numerical validation of proposed fusion technique

Experimental results and numerical validation of the proposed data fusion measurement technique and sensitivity analysis of various γ values ($\gamma \approx 0.01, 0.1, 0.5, 1, 10, 100$) are reported in Figure 4. As γ approaches zero (i.e., $\gamma \ll 1$), the fused feature distribution approximates the marginal contribution of the higher-dimensional source. As γ approaches infinity (e.g., $\gamma \gg 1$), the fused feature distribution approximates the lower-dimensional source. That is, when γ is extremely small or large (i.e., $\gamma \approx 0.01$ or $\gamma \approx 100$), the fused distribution approximates one of the source distributions in feature space. In either case, the fused distribution is insufficient to convey information of *both* source domains. When $\gamma \approx 1$, the probability density function of the fused data demonstrates a balanced shape between the two source functions, implying that neither source domain has a dominating effect on the fused feature space. The energy differences shown

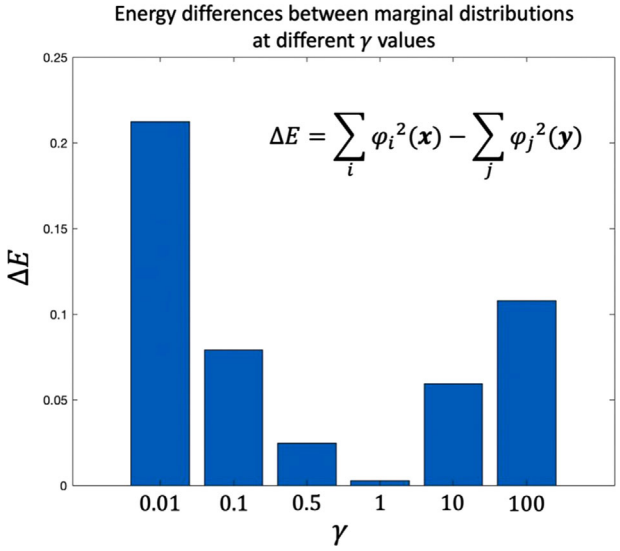

FIGURE 5 Energy differences at different γ values. The energy difference between the marginal distribution of the Gaussian process and the sparsely sampled points indicate differences in fusion quality. Optimal fusion is expected at $\gamma = 1$, where the energy is minimized between the two domain-specific state functions.

in Figure 5 quantify this effect of fusion quality. As the fused distribution becomes biased towards either source domain, the energy difference between the two domains increases monotonically. Energy is minimized when $\gamma \approx 1$, indicating a optimal fusion quality between the two source domains.

3.2 | Deep fusion of CT imaging and hepatic blood markers to estimate portal venous hypertension

Overall model performance is reported in Table 1. The joint model consisting of both imaging data and positionally encoded blood markers at $\gamma \approx 0.99$ achieved an AUC of 0.74 and an accuracy of 0.71, resulting in an increase in model performance relative to the imaging-only model (AUC = 0.60; accuracy = 0.62) and the blood markers-only model (AUC = 0.58; accuracy = 0.60). The joint model at $\gamma \approx 0.99$ had the best test AUC, accuracy, and the best ability to generalize, which suggests that CT imaging and blood markers provide complementary diagnostic information, but only if appropriately fused.

As an illustrating example, Figure 6 reports measured feature density functions at monotonically increasing d_{2*} values from a representative case. In the absence of positional encoding, the fused feature distribution closely approximates the distribution of imaging features. When γ is much smaller than 1, i.e., $\gamma \approx 0.09, 0.19, 0.38$, the fused distribution is still closer to the imaging distribution and cannot well represent the blood marker features. As γ increases, the joint distribution gradually approaches the distribution of blood marker features. When $\gamma \approx 0.99$, the fused distribution takes on a shape that neither represents imaging or blood markers, indicating an optimal fusion of the two data domains. As γ grows larger, the fused distribution begins to approximate the positionally encoded blood marker distribution.

TABLE 1 Performance comparison of different modeling techniques on testing set

	Test AUC	Test Accuracy
Image Only	0.60	0.62
Blood Markers Only	0.59	0.60
$\gamma \approx 0.99$ **PE Blood Markers**	0.58	0.60
Image + No PE Blood Markers	0.60	0.62
Image $+\gamma \approx 0.09$ **PE Blood Markers**	0.64	0.60
Image $+\gamma \approx 0.19$ **PE Blood Markers**	0.61	0.58
Image $+\gamma \approx 0.38$ **PE Blood Markers**	0.61	0.61
Image $+\gamma \approx 0.99$ **PE Blood Markers**	0.74	0.71
Image $+\gamma \approx 1.50$ **PE Blood Markers**	0.70	0.69
Image $+\gamma \approx 3.00$ **PE Blood Markers**	0.63	0.57

As demonstrated in Figure 7, model performance results were associated with intrinsic differences in feature density estimation. Consistent with theory, model performance peaked when fused at $\gamma \approx 0.99$, which resulted in a fused density distribution different from both source distributions.

4 | DISCUSSION

Deep neural networks play an increasingly important role in the staging,[30] detection,[31] characterization,[32] segmentation,[33] classification,[34] and computer-aided diagnosis[35–41] of disease on imaging. However, optimal integration with clinical data – which is often sparsely encoded and low-dimensional – requires sophisticated data fusion techniques. In this work, we developed a new data fusion technique and quality metric for deep learning problems involving both imaging data and clinical data. Our approach is motivated by the mathematical methods of statistical mechanics, which play an increasingly important role in complex data modeling[42] that complements mainstream AI techniques and has demonstrated utility in imaging radiomics problems.[43–45] We then demonstrated the feasibility of our approach by estimating portal venous hypertension based on the deep fusion of CT imaging and hepatic blood markers. While this illustrating example was chosen due to its clinical relevance and interest to the authors, the proposed data fusion formalism is general and can therefore be extended to other deep learning problems in radiology.

Prior studies have investigated the effect of augmenting deep image representation with non-imaging,

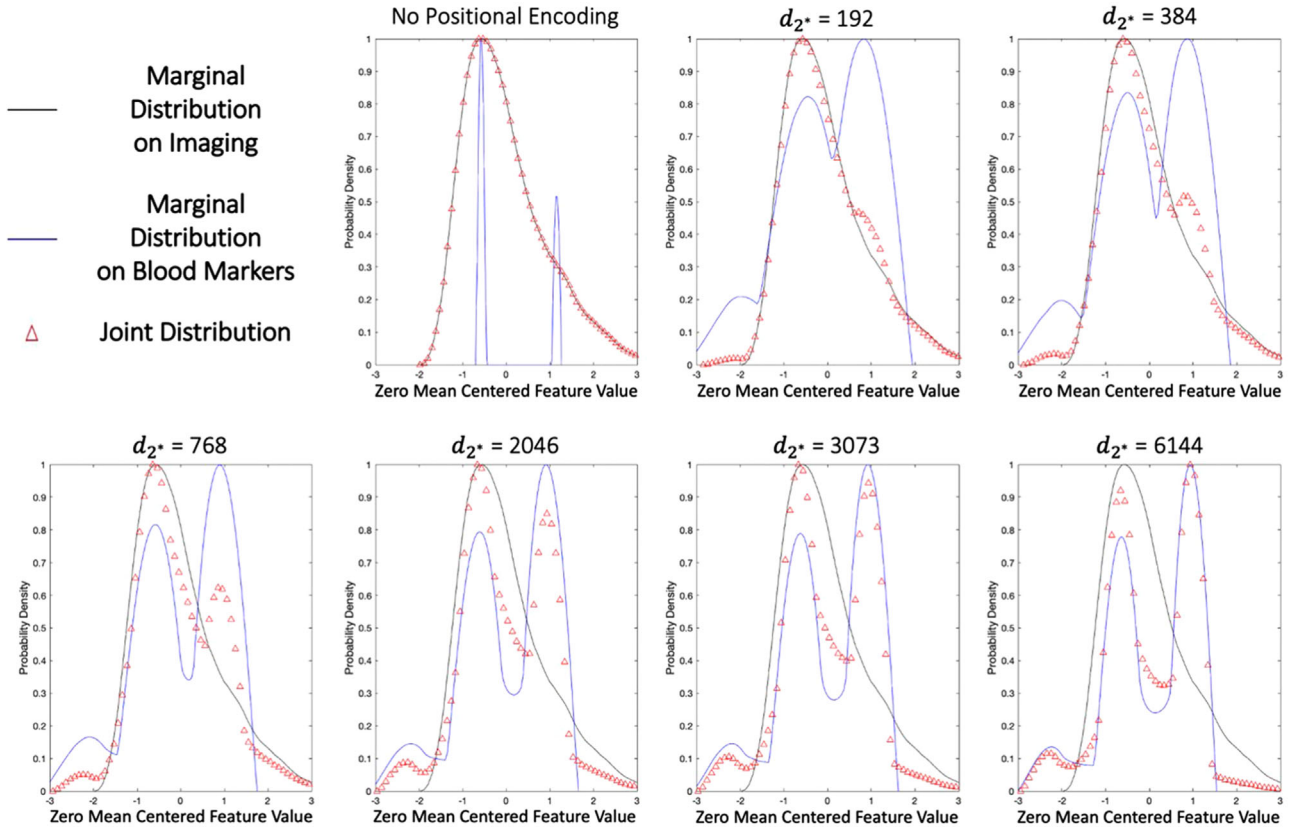

FIGURE 6 The marginal distribution on the real ResNet imaging feature vector, marginal distribution of the blood marker feature vector, and their joint distribution as measured on a representative patient case for various monotonically increasing values of d_{2^*} (d_{2^*} = 192, 384, 768, 2046, 3073, 6144).

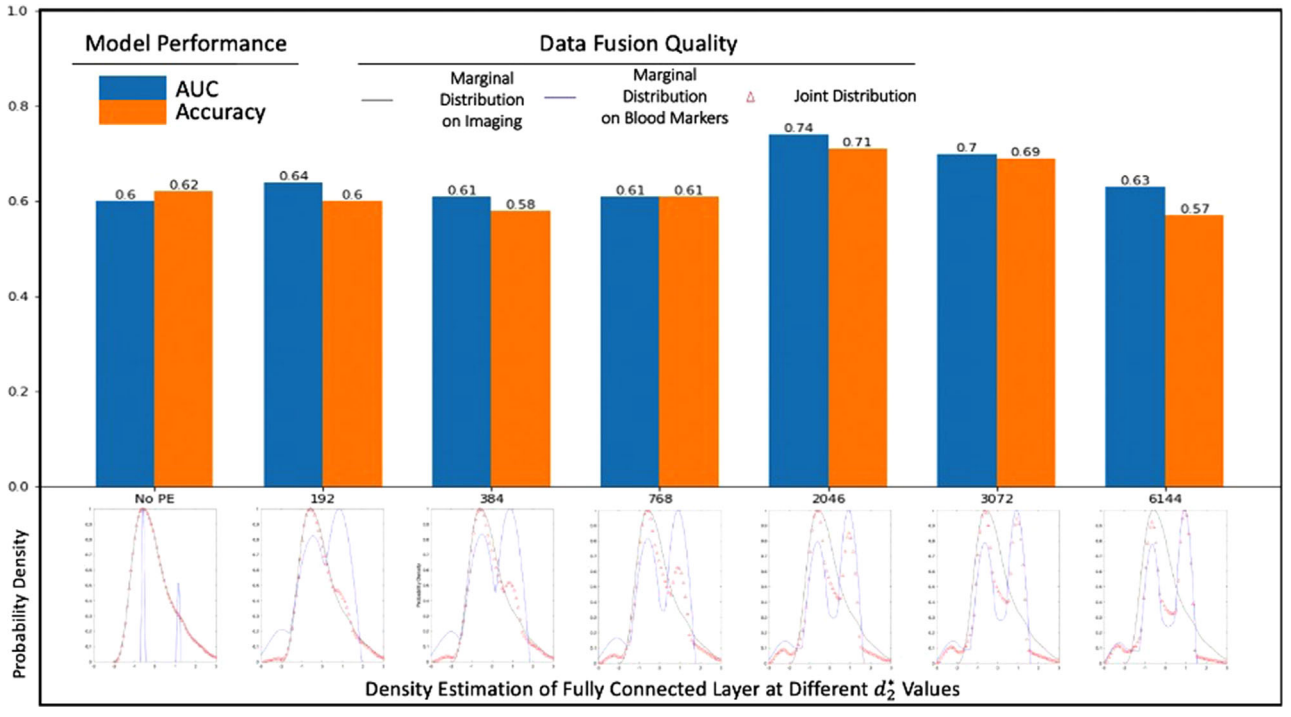

FIGURE 7 Model performance and feature density visualization as a function of d_{2^*}. (Top) The AUC and accuracy of the model is shown at different values of d_{2^*}. (Bottom) The measured distribution on the ResNet imaging feature vector, measured distribution on the blood marker feature vector, and their measured joint distribution is shown from a representative case at different values of d_{2^*}. Differences in fused feature density are associated with measurable changes in model performance, which peaks at the expected d_{2^*} = 2046

clinically relevant data.[6–15] Many of these papers reported improved model performance when incorporating multiple sources of data,[6–12,15] thus demonstrating the benefits of multi-source modeling. However, a key and common limitation is their lack of rigorous data fusion implementation. This is important because the performance and reproducibility of multi-source deep learning models are sensitive to *how* the data is fused.[16] For example, Yoo et al. reported that joint fusion was better than late fusion of deep MRI features and patient data when predicting multiple sclerosis.[9] Their work illustrates that improper data fusion may discard otherwise useful information.

As such, a purely heuristic approach to data fusion may lead to sub-optimal applicability. For example, Yap et al. concatenated raw patient data directly (i.e., without any pre-processing or positional encoding scheme) with the FCL of their ResNet architecture.[10] The performance of their fused model did not improve relative to the image-only model, which is consistent with our results. We hypothesize that this approach is analogous to adding noise to the ResNet embedding, because in this scenario, the patient features are sparsely encoded and therefore dominated by the higher-dimensional image feature space. Our proposed approach to model the features of the FCL as random variables governed by source-specific state functions provides a means to directly measure fusion quality and identify source bias.

Our results suggest that CT imaging and hepatic blood markers can provide complementary information, but only if they are appropriately fused within the deep learning architecture. We observed comparable model performance when considering imaging data and blood marker data independently. Fusion of these data into a single model *without* positional encoding of the blood markers resulted in the same exact performance as the image-only model. This implies a source bias of the image features, which was the dominating effect of the joint model.

When we implemented positional encoding of the blood marker data, model performance peaked when the energy difference between the two data sources was minimized according to Equation (5). This finding is consistent with our theoretical formalism, as well as our numerical analysis on simulated Gaussian processes. In this low-energy state where $\gamma \approx 1$, the fusion of imaging data and blood marker data is optimized according to the proposed theory. Uncoincidentally, this theoretically optimal parameterization also demonstrated the highest downstream model performance.

Model performance decreased as the d_{2*} value of the blood markers was increased beyond the dimension of the ResNet imaging features. In this scenario, in the limit that $d_{2*} \gg d_1$, the consistency condition is not satisfied according to Equations (6) and (7). Coincidentally, the blood marker data became the dominating effect of the network's soft max operation, which adversely affected model performance. These findings are again consistent with theory and numerical results.

This paper therefore provides a new way to measure fusion quality in deep learning imaging problems, verifies the approach based on numerical simulation, and demonstrates feasibility on clinically relevant, real-world data. However, our work is not without limitation. First, the fusion quality metric derived in this paper is the special solution of $n = 2$ source domains (e.g., imaging data + blood marker data). In future work, we plan to study the more complex, general solution of $n > 2$ source domains and cases where the optimal point is not achieved when data sources are equally weighted. Second, our experimental design was based on a retrospective dataset limited in sample size due to relevant inclusion criteria. While our experimental results were largely consistent with our theoretical formulation and numerical analysis, future application work should be based on a larger dataset. Vector growing in particular may have implications to model overfitting, which should be carefully investigated on a larger dataset. Third, we chose to implement a positional encoding scheme based on its simplicity and its popularity in other fields. Our goal here was to increase the dimension of blood markers and avoid changing its contained information, which is the relative size of each data point. Intuitively, the relative position of a data point among all possible values demonstrates the same information as its relative size. However, there are a variety of other encoding methods (e.g., linear projection, principal component analysis) that can be investigated as future work, all of which can be studied via our fusion quality approach. Finally, we modified the ResNet-152 model to extract imaging features and make the predictions at the same time. This design choice was based on the well-known effectiveness of ResNet-152 as an image encoder. However, other predictors such as multi-layer FCL with a better performance in non-linear problems are also worth exploring as future work, which can be coupled with our proposed data fusion technique.

5 | CONCLUSIONS

In this work, we introduced the concept of data fusion quality for multi-source deep learning problems. We provided a rigorous theoretical framework, numerical validation, and real-world application in abdominal radiology. Our data suggests that CT imaging and hepatic blood markers provide complementary diagnostic information when appropriately fused. This is a clinically relevant finding, because there is growing scientific evidence that portal venous hypertension is an important biomarker to identify patients at risk for cirrhosis. The mathematical formalism proposed in this paper can be applied to other applications in diagnostic medical imaging.

ACKNOWLEDGMENT

None.

CONFLICT OF INTEREST

None

REFERENCES

1. JDL, Data Fusion Lexicon. *Technical Panel For C3*, F.E. White, Code 420. 1991.
2. Hall DL, Llinas J. An introduction to multisensor data fusion, *in Proc. IEEE*. 1997;85(1):6-23. https://doi.org/10.1109/5.554205
3. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med*. 2020;3:136. Published 2020 Oct 16. https://doi.org/10.1038/s41746-020-00341-z
4. Kharazmi P, Kalia S, Lui H, Wang ZJ, Lee TK. A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile. *Skin Res Technol*. 2018;24(2):256-264. https://doi.org/10.1111/srt.12422
5. Li H, Fan Y. Alzheimer's disease neuroimaging initiative. Early prediction of Alzheimer's disease dementia based on baseline hippocampal MRI and 1-year follow-up cognitive measures using deep recurrent neural networks. *Proc IEEE Int Symp Biomed Imaging*. 2019;2019:368-371. https://doi.org/10.1109/ISBI.2019.8759397
6. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*. 2019;292(1):60-66. https://doi.org/10.1148/radiol.2019182716
7. Yoo Y, Tang LYW, Li DKB, et al. Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome, *Comput Methods Biomech Biomed Eng Imaging Vis*. 2019;7(3):250-259. doi: 10.1080/21681163.2017.1356750
8. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol*. 2018;27(11):1261-1267. https://doi.org/10.1111/exd.13777
9. Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM. Alzheimer's disease neuroimaging initiative. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput Biol*. 2018;14(9):e1006376. Published 2018 Sep 14. https://doi.org/10.1371/journal.pcbi.1006376
10. Purwar S, Tripathi RK, Ranjan R, Saxena R. Detection of microcytic hypochromia using cbc and blood film features extracted from convolution neural network by different classifiers. *Multim Tools Appl Dordrecht*. 2020;79(7-8):4573-4595. https://doi.org/10.1007/s11042-019-07927-0
11. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-Point checklist and skin lesion classification using multi-task multimodal neural nets [published online ahead of print, 2018 Apr 9]. *IEEE J Biomed Health Inform*. 2018. https://doi.org/10.1109/JBHI.2018.2824327
12. Qiu S, Chang GH, Panagia M, Gopal DM, Au R, Kolachalama VB. Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimers Dement (Amst)*. 2018;10:737-749. Published 2018 Sep 28. https://doi.org/10.1016/j.dadm.2018.08.013
13. Reda I, Khalil A, Elmogy M, et al. Deep learning role in early diagnosis of prostate cancer. *Technol Cancer Res Treat*. 2018;17:1533034618775530. https://doi.org/10.1177/1533034618775530
14. Rothe S, Kudszus B, Söffker D. Does classifier fusion improve the overall performance? Numerical analysis of data and fusion method characteristics influencing classifier fusion performance. *Entropy (Basel)*. 2019;21(9):866. Published 2019 Sep 5. https://doi.org/10.3390/e21090866
15. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need, *Adv. Neural Sci.* Information Processing Systems. 2019; pp. 5998–6008. doi: 10.48550/arXiv.1706.03762
16. Gibbs, JW (1902). *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons
17. Landau, LD & Lifshitz, EM (1980) [1976]. *Statistical Physics. Course of Theoretical Physics*. Vol. 5 (3rd ed.). Pergamon Press. ISBN 0-7506-3372-7. Translated by J.B. Sykes and M.J. Kearsley.
18. Boltzmann, L (1868). "Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten" [Studies on the balance of living force between moving material points]. *Wiener Berichte*. 58:517–560.
19. Magna AAR, Allende-Cid H, Taramasco C, Becerra C, Figueroa RL, Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis. *IEEE Access*. 2020; 8:106198-106213. https://doi.org/10.1109/ACCESS.2020.3000075
20. Si S, Wang R, Wosik J, et al. Students need more attention: bert-based attentionmodel for small data with application to automatic patient message triage. *Mach Learn Healthcare*. 2020. doi: 10.48550/arXiv.2006.11991
21. Liu S, Yadav C, Fernandez-Granda C, Razavian N. On the design of convolutional neural networks for automatic detection of Alzheimer's disease. *NeurIPS ML4H*. 2019. doi: 10.48550/arXiv.1911.03740
22. Heidelbaugh JJ, Sherbondy M. Cirrhosis and chronic liver failure: part II. Complications and treatment. *Am Fam Physician*. 2006;74(5):767-776. url: https://www.aafp.org/afp/2006/0901/afp20060901p767.pdf
23. Tseng Y, Ma L, Luo T, et al. Non-invasive predictive model for hepatic venous pressure gradient based on a 3-dimensional computed tomography volume rendering technology. *Exp Ther Med*. 2018;15(4):3329-3335. https://doi.org/10.3892/etm.2018.5816
24. Smith AD, Branch CR, Zand K, et al. Liver surface nodularity quantification from routine CT images as a biomarker for detection and evaluation of cirrhosis [published correction appears in Radiology. 2017 Jun;283(3):923]. *Radiology*. 2016;280(3):771-781. https://doi.org/10.1148/radiol.2016151542
25. Sartoris R, Rautou PE, Elkrief L, et al. Quantification of liver surface nodularity at CT: utility for detection of portal hypertension. *Radiology*. 2018;289(3):698-707. https://doi.org/10.1148/radiol.2018181131
26. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. https://doi.org/10.1038/s41592-020-01008-z
27. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv*. 2015; doi: 10.48550/arXiv.1505.04597
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 770–778, https://doi.org/10.1109/CVPR.2016.90
29. He F, Liu T, and Tao D. Why ResNet works? Residuals generalize. *IEEE Trans Neural Netw Learn Syst*. 2020;31(12):5349-5362. https://doi.org/10.1109/TNNLS.2020.2966319
30. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images. *Radiology*. 2018;287(1):146-155. https://doi.org/10.1148/radiol.2017171928
31. Vivanti R, Szeskin A, Lev-Cohain N, Sosna J, Joskowicz L. Automatic detection of new tumors and tumor burden evaluation in longitudinal liver CT scan studies. *Int J Comput*

Assist Radiol Surg. 2017;12(11):1945-1957. https://doi.org/10.1007/s11548-017-1660-z
32. Biswas M, Kuppili V, Edla DR, et al. Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Comput Methods Programs Biomed*. 2018;155:165-177. https://doi.org/10.1016/j.cmpb.2017.12.016
33. Lu F, Wu F, Hu P, Peng Z, Kong D. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. *Int J Comput Assist Radiol Surg*. 2017;12(2):171-182. https://doi.org/10.1007/s11548-016-1467-3
34. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of deep learning approaches for multi-label chest X-ray classification. *Sci Rep*. 2019;9(1):6381. Published 2019 Apr 23. https://doi.org/10.1038/s41598-019-42294-8
35. Liu Y, Ning Z, Örmeci N, et al. Deep convolutional neural network-aided detection of portal hypertension in patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2020;18(13):2998-3007.e5. https://doi.org/10.1016/j.cgh.2020.03.034
36. Draelos RL, Dov D, Mazurowski MA, et al. Machine-learning-based multiple abnormality prediction with large-scale chest computed tomography volumes. *Med Image Anal*. 2021;67:101857. https://doi.org/10.1016/j.media.2020.101857
37. Li Q, Zhang Y, Liang H, et al. Deep learning based neuronal soma detection and counting for Alzheimer's disease analysis. *Comput Methods Programs Biomed*. 2021;203:106023. https://doi.org/10.1016/j.cmpb.2021.106023
38. El Adoui M, Drisis S, Benjelloun M. Multi-input deep learning architecture for predicting breast tumor response to chemotherapy using quantitative MR images. *Int J Comput Assist Radiol Surg*. 2020;15(9):1491-1500. https://doi.org/10.1007/s11548-020-02209-9
39. Koshimizu H, Kojima R, Kario K, Okuno Y. Prediction of blood pressure variability using deep neural networks. *Int J Med Inform*. 2020;136:104067. https://doi.org/10.1016/j.ijmedinf.2019.104067
40. Liu X, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors (Basel)*. 2017;17(1):149. Published 2017 Jan 13. https://doi.org/10.3390/s17010149
41. Ben-Cohen A, Klang E, Diamant I, et al. CT image-based decision support system for categorization of liver metastases into primary cancer sites: initial results. *Acad Radiol*. 2017;24(12):1501-1509. https://doi.org/10.1016/j.acra.2017.06.008
42. Lafata KJ, Zhou Z, Liu JG, Yin FF. Data clustering based on Langevin annealing with a self-consistent potential. *Q Appl Math*. 2019 Sep;77(3):591-613.
43. Lafata KJ, Chang Y, Wang C, et al. Intrinsic radiomic expression patterns after 20 Gy demonstrate early metabolic response of oropharyngeal cancers. *Med Phys*. 2021 Jul;48(7):3767-3777.
44. Lafata KJ, Corradetti M, Gao J, et al. Radiogenomic analysis of locally advanced lung cancer based on CT imaging and intra-treatment changes in cell free DNA. *Radiol Imaging Cancer*. 2021 Apr;3(4):e200157.
45. Lafata KJ, Zhou Z, Liu JG, Hong J, Kelsey C, Yin FF. An exploratory radiomics approach to quantifying pulmonary function in CT images. *Sci Rep*. 2019 Aug;9:11509.

How to cite this article: Wang Y, Li X, Konanur M, et al. Towards optimal deep fusion of imaging and clinical data via a model-based description of fusion quality. *Med Phys*. 2023;1-12. https://doi.org/10.1002/mp.16181