Received: 8 January 2025 | Revised: 27 March 2025 | Accepted: 9 April 2025

DOI: 10.1002/mrm.30548

RESEARCH ARTICLE

Magnetic Resonance in Medicine

CEST MRI data analysis using Kolmogorov-Arnold network (KAN) and Lorentzian-KAN (LKAN) models

Jiawen Wang | **Pei Cai** | **Ziyan Wang** | **Huabin Zhang** | **Jianpan Huang**

Laboratory of Advanced Imaging in Medicine (AIM), Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

Correspondence
Jianpan Huang, Laboratory of Advanced Imaging in Medicine (AIM), Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 5 Sassoon Road, Pokfulam, 999077 Hong Kong, China.
Email: jphuang@hku.hk

Funding information
National Natural Science Foundation of China, Grant/Award Number: 82402225; University of Hong Kong, Grant/Award Numbers: 109000487, 109001694, 204610401, 204610519

Abstract

Purpose: To investigate the potential of using Kolmogorov-Arnold Network (KAN) and propose Lorentzian-KAN (LKAN) for CEST MRI data analysis (CEST-KAN/CEST-LKAN).

Methods: CEST MRI data acquired from 27 healthy volunteers at 3 T were used in this study. Data from 25 subjects were used for training and validation (548 865 Z-spectra), whereas the remaining two were reserved for testing (51 977 Z-spectra). The performance of multi-layer perceptron (MLP), KAN, and LKAN models was evaluated and compared to conventional multi-pool Lorentzian fitting (MPLF) method in generating ΔB_0, water, and multiple CEST contrasts, including amide, relayed nuclear Overhauser effect (rNOE), and magnetization transfer (MT).

Results: The KAN and LKAN showed higher accuracy in predicting CEST parameters compared to MLP, with average reductions in test loss of 28.37% and 32.17%, respectively. Voxel-wise correlation analysis also revealed that ΔB_0 and four other CEST parameters from the KAN and LKAN had higher average Pearson coefficients than MLP by 1.57% and 2.84%, indicating superior performance. LKAN exhibited a shorter average training time by 37.26% and a smaller average test loss by 5.29% compared to the KAN. Furthermore, our results demonstrated that even smaller KAN and LKAN could achieve better accuracy than MLPs, with both KAN and LKAN showing greater robustness to noisy data compared to MLP.

Conclusion: This study demonstrates the feasibility of KAN and LKAN for CEST MRI data analysis, highlighting their superiority over MLP. The findings suggest that CEST-KAN and CEST-LKAN have the potential to be robust and reliable post-analysis tools for CEST MRI in clinical settings.

KEYWORDS

chemical exchange saturation transfer (CEST), Kolmogorov-Arnold network (KAN), Lorentzian-KAN (LKAN), multi-pool Lorentzian fitting (MPLF), human brain

Jiawen Wang and Pei Cai contributed equally to this work.

1 | INTRODUCTION

CEST MRI is a molecular imaging technique that enables the detection of low-concentration molecules through the exchange between water protons and exchangeable solute protons.[1–4] It has been extensively used for imaging various endogenous contrasts, including amide,[5–8] glutamate,[9–12] creatine,[13–17] and glucose.[18–23] Additionally, apart from the direct CEST effects, the relayed nuclear Overhauser effects (rNOE) at negative frequency offsets on the water saturation spectrum (Z-spectrum) have gained significant attention over the past years.[24–29] By leveraging the in vivo biomolecules or compounds with exchangeable protons, CEST MRI holds immense potential for disease diagnosis and treatment monitoring.

Despite its potential for clinical use, there are several limitations that hinder the widespread application of CEST MRI. One of these limitations is the absence of a standardized and universally accepted postprocessing method for CEST MRI. Moreover, most existing postprocessing methods are often time-consuming, complex, and require specialized knowledge and expertise. The existing methods include magnetization transfer (MT) ratio asymmetry (MTR_{asym}),[6,30] Lorentzian difference analysis (LDA),[24,31] multi-pool Lorentzian fitting (MPLF),[32,33] polynomial and Lorentzian line-shape fitting (PLOF),[17,34] three-point method[28,35] and other methods.[36,37] Among these methods, MPLF is a frequently used CEST postprocessing method.[33] This method typically fits the Z-spectra acquired under low saturation powers by assuming a Lorentzian line shape for each exchanging proton pool. The amplitude obtained from the fit is then used to quantify the signal of the corresponding exchanging proton pool. MPLF generally works well, especially at high field strengths where the individual CEST peaks can be well distinguished, but its time-consuming nature poses a challenge to its widespread clinical implementation for 3D human CEST data.

To accelerate and simplify the CEST postprocessing, many deep learning techniques, mostly multi-layer perceptron (MLP)-based models, have been extensively used in extracting various CEST fitting parameters from Z-spectra with faster speed than conventional methods in recent years.[38–44] MLP-based models have also been used in CEST magnetic resonance fingerprinting (CEST-MRF), which is a quantitative CEST technique, to speed up the reconstruction process.[45–51] Encouraging results from these studies demonstrated the significant promise of leveraging MLP techniques for CEST analysis. Nevertheless, despite the widespread adoption of MLPs, they exhibit notable drawbacks. For instance, MLPs are less interpretable without the aid of post-analysis tools[52,53] and are sensitive to noise in node features, which can compromise their performance.[54] Recently, the Kolmogorov-Arnold Network (KAN)[53] has been proposed as promising alternatives to MLPs. Unlike MLPs with fixed activation functions, KANs replace traditional linear weight matrices with learnable activation functions at each weight parameter. This innovative approach allows KANs to provide smaller yet more accurate models with faster scaling laws and enhanced interpretability.[53]

In this study, we investigated the potential of KAN for analyzing CEST MRI data for the first time. Furthermore, we propose a novel KAN-based network, termed Lorentzian-KANs (LKANs), which incorporates Lorentzian functions that mathematically align with the shapes of CEST peaks extracted using MPLF. This physics-informed modification is expected to improve both training efficiency and accuracy, while also enhancing the interpretability of the model. We rigorously evaluated the performance of the MLP, KAN, and LKAN models for predicting the MPLF parameters on the same human CEST dataset. Additionally, we investigated the influence of various hyperparameters on the model's performance. We also compared the predicted CEST fitting parameters of the MLP, KAN, and LKAN models across different noise levels and explored the interpretability of these models.

2 | METHODS

2.1 | Theory

2.1.1 | MPLF

In CEST postprocessing, Z-spectra are generated by normalizing the CEST source images $M_{sat}(\Delta\omega)$ to the M_0 image:

$$Z(\Delta\omega) = \frac{M_{sat}(\Delta\omega)}{M_0}, \tag{1}$$

where $\Delta\omega$ is the saturation frequency offsets of Z-spectra. In MPLF analysis, the CEST peak of each exchange pool (i) can be depicted by a Lorentzian line shape[32]:

$$L_i(\Delta\omega) = \frac{A_i}{1 + \left[\frac{\Delta\omega - (\Delta\delta_i + \Delta\delta)}{\Gamma_i/2}\right]^2}, \tag{2}$$

where A_i, $\Delta\delta_i$, and Γ_i represent the peak amplitude, peak position, and full-width-at-half-maximum (FWHM) of exchange pool i, $\Delta\delta$ is the B_0 inhomogeneity in ppm. Therefore, the Z-spectra can be fitted by the MPLF method using the following equation:

$$Z_{fit}(\Delta\omega) = 1 - \sum_{i}^{N} L_i(\Delta\omega), \tag{3}$$

where N is the total number of exchange pools. Considering the CEST data used in this study were acquired from a clinical 3 T MRI scanner where the spectral resolution is relatively low, we used a four-pool Lorentzian fitting model, which considers direct water saturation (DS) at 0 ppm, amide at 3.5 ppm, rNOE at −3.5 ppm and MT at −2.5 ppm. MPLF analysis was performed using custom-written code in MATLAB (The MathWorks). The initial values and bound conditions for MPLF are given in Table S1.

2.1.2 | MLPs

MLPs refer to fully connected feedforward neural networks that are the foundational blocks of various deep learning models.[55–57] MLPs are based on the universal approximation theorem structuring computations through layered transformations, which can be simplified as:

$$MLP(x) = (W_{L-1} \circ \sigma \circ W_{L-2} \circ \sigma \circ \cdots \circ W_1 \circ \sigma \circ W_0)x, \tag{4}$$

where x denotes the input (Z values in this study), L refers to the total number of layers of MLP, W and σ represent the linear transformations and non-linear activation functions (such as sigmoid or ReLU), respectively. In MLP (Figure 1), each node represents a neuron, and each connection between the nodes corresponds to a learnable layer weight.

2.1.3 | KANs

KANs are novel fully connected deep neural networks introduced with a radical shift from the MLP paradigm by leveraging the Kolmogorov-Arnold theorem. This theorem enables the representation of any multivariate continuous function as a sum of univariate functions. In KANs, the input (denoted as x, which refers to Z values in this study) is processed across layers. Denoting the activation function that connects i^{th} neuron in l^{th} layer and j^{th} neuron in $l+1^{th}$ layer as $\phi_{l,j,i}$, KANs can be expressed as:

$$KAN(x) = \sum_{i_{L-1}=1}^{n_{L-1}} \phi_{L-1,i_L,i_{L-1}} \times \left(\sum_{i_{L-2}=1}^{n_{L-2}} \cdots \left(\sum_{i_1=1}^{n_1} \phi_{1,i_2,i_1} \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0}(x_{i_0}) \right) \right) \cdots \right), \tag{5}$$

where each $\phi_{l,j,i}$ is the sum of the spline function and the basis function $b(x)$, serving as residual connections:

$$\phi(x) = \omega_b b(x) + \omega_s spline(x), \tag{6}$$

$$spline(x) = \sum_i c_i B_i(x). \tag{7}$$

For better understanding, a general KAN of L layers can be simplified as:

$$KAN(x) = (\phi_{L-1} \circ \phi_{L-2} \cdots \phi_1 \circ \phi_0)x, \tag{8}$$

where x denotes the input (Z values in this study), L refers to the total number of layers of KANs, and ϕ_l is the function matrix corresponding to the l^{th} layer.

Unlike MLPs, which treat linear transformations and nonlinearities separately as W and σ, KANs integrate both together in ϕ by summing incoming signals transformed by a set of B-spline functions. As a result, no explicit linear weight matrices are used. Instead, KANs use learnable activation functions parameterized as B-splines on the edges of each layer. Intuitively speaking, the summation of these activation functions suggests that the KANs are combinations of MLPs and B-splines—exhibiting an MLP shape externally while using a spline design internally. Consequently, this design leverages the accuracy of B-splines for low-dimensional functions while mitigating the curse of dimensionality because of its external resemblance to MLPs.[53]

2.1.4 | LKANs

LKANs are a novel neural network architecture proposed in this study for analyzing CEST MRI data. While KANs traditionally use B-splines to approximate functions, LKANs innovate by replacing them with Lorentzian functions:

$$\psi(x) = \omega_b b(x) + \omega_l Lorentz(x), \tag{9}$$

$$Lorentz(x) = \sum_i c_i L(x), \tag{10}$$

$$L(x) = \frac{\gamma^2}{(x - x_0)^2 + \gamma^2}. \tag{11}$$

This modification leverages the unique properties of Lorentzian functions, such as broader bilateral tails and sharper peaks, to capture both local and global patterns in CEST data more effectively. By incorporating Lorentzian functions, LKANs retain the adaptive qualities of KANs while enhancing their ability to model CEST Z-spectra, which are characterized by Lorentzian line shapes when using MPLF as the analysis method. Therefore, the proposed LKANs can be consistently simplified as:

$$LKAN(x) = (\psi_{L-1} \circ \psi_{L-2} \cdots \psi_1 \circ \psi_0)x$$

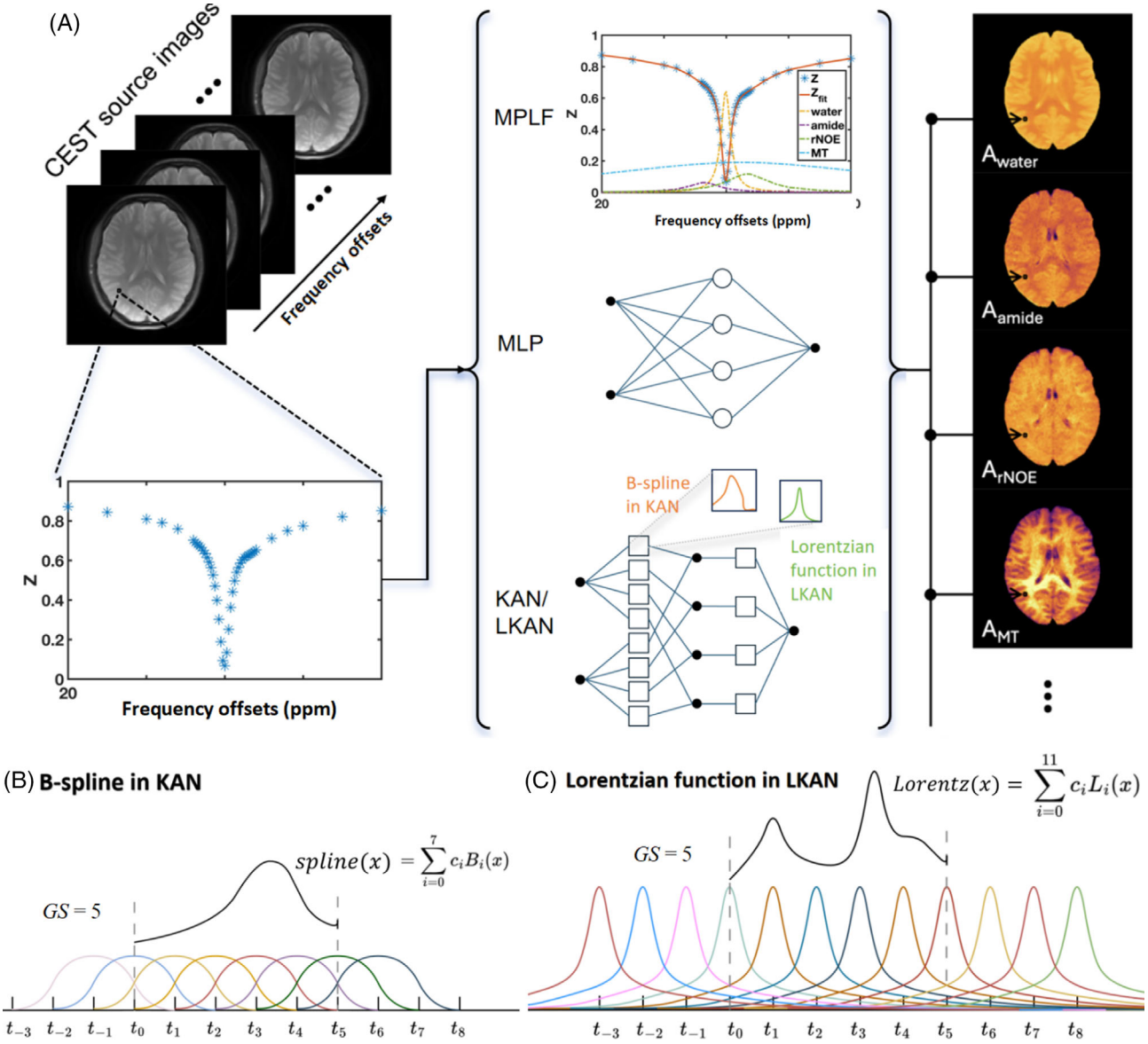

FIGURE 1 (A) Illustration of data analysis for CEST MRI using multi-pool Lorentzian fitting (MPLF), multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN) methods. The input is the Z-spectrum, and the output is multiple CEST fitting parameters. (B) Schematic illustration of the activation function of KAN parameterized by B-splines with grid size = 5. (C) Schematic illustration of the activation function of LKAN parameterized by Lorentzian functions with grid size = 5.

where x denotes the input (Z values in this study), L refers to the total number of layers of LKANs, and ψ_l is the Lorentzian function matrix corresponding to the l^{th} layer. Therefore, LKANs build on the conceptual framework of KANs, maintaining their core structure while exhibiting enhanced performance in contexts where the data displays specific physics-informed characteristics.

In the practical implementation of LKANs and KANs, Lorentzian and B-spline functions, despite both being radial basis functions, exhibit notable differences. The effective range of the B-spline function is limited by the spline order, whereas the Lorentzian function maintains non-zero values across the real number domain. As shown in Figure 1, the number of B(x) functions equals the grid size (GS) plus the spline order. Additionally, the interval is extended beyond the original boundaries by a margin equal to the spline order multiplied by the GS on both sides. This concept is akin to the padding operation in convolutional networks. In LKANs, each Lorentzian curve is centered at a grid point, with the effective interval extended by P grid units on both sides. Here, we define P as the padding factor. This approach allows us to approximate Lorentzian curves that are centered more than P grid sizes away from the interval boundaries as zero.

2.2 | MRI scan

The study was approved by the institutional review board of The University of Hong Kong and conducted according to the guidelines. Single-slice (2D) CEST MRI experiments were performed at a 3 T MRI system (SIGNA Premier, GE Healthcare) on 30 voluntary participants with an average age of 42.76 years old (15 males and 15 females, age 22–74 years old). Individuals whose head motion showed translations exceeding ±2 mm or rotations exceeding ±2 degrees in any directions assessed by SPM12 were excluded. Additionally, incidental findings of any intracranial pathologies were also considered as exclusion criteria. Ultimately, CEST data from 27 healthy volunteers with an average age of 41.40 years (13 males and 14 females, age 22–73 years) were used for further analysis. The CEST sequence comprised a continuous wave (CW) module with a saturation time of 2 s and a saturation power of 0.8 μT, and a single-shot fast spin echo (FSE) readout module. M_0 images at a frequency offset of −300 ppm and 43 CEST images at frequency offsets ranging from −20 to 20 ppm were acquired. Other parameters were as follows: TR = 3 s, TE = 23 ms, FOV = 220 × 220 mm^2, acquisition matrix = 128 × 128, reconstructed matrix size = 282 × 282, slice thickness = 6 mm. The scan time for each CEST data was 2 min 21 s.

2.3 | Network training

For CEST-MLP, CEST-KAN, and CEST-LKAN, the input of networks was the full Z-spectrum of each image voxel in a length of 43, and the output of networks was the nine CEST fitting parameters derived from the MPLF method, with the pool number set to four, as illustrated in Figure 1A. Instead of using the conventional method of acquiring a B_0 map and performing an interpolation-based B_0 correction, we incorporated the B_0 inhomogeneity directly into the fit model (Eq. [2]). The training and validation of the networks used CEST data from 25 participants, comprising 548 865 Z-spectra, with 80% (439 092 Z-spectra) used for training and 20% (109 773 Z-spectra) for validation. The remaining CEST data from two participants, comprising 51 977 Z-spectra, was reserved for testing.

The networks were implemented using PyTorch with the following initial configurations: GS in KAN = 9, spline order in KAN = 3, GS in LKAN = 1, P in LKAN = 5. During network training, a batch size of 32 was used, and the learning rate was set to 0.01 and decayed with a factor of 0.8 every epoch. The optimizer used was AdamW,[58] a variant of the Adam optimizer with a weight decay (set to 10^{-4} in this study). The loss function used was mean square error (MSE). The early stop patience was set to 50. The network training was conducted on a workstation (Intel Core i9, 64 GB memory) with an NVIDIA RTX 4090 graphics processing unit (GPU). To ensure the reliability of the results, each model in this study was trained for five times, and the average results were used for comparison.

2.3.1 | Hyperparameter exploration

GS for KAN

To evaluate the influence of GS on the performance of KAN, we systematically varied the GS from 1 to 10 in increments of 1 on one-layer KAN models (shape [43, 100, 9]). The results comparing the validation loss and the number of trainable parameters for each network are presented in Section 3.1.1.

GS and P for LKAN

To investigate the impact of GS and P on the performance of LKAN, we set the GS to 1, 3, and 5, while adjusting the P in LKAN to 1, 3, 5, 7, and 9 on one-layer LKAN models (shape [43, 100, 9]). The results comparing the validation loss and the number of trainable parameters for each network are presented in Section 3.1.2.

2.3.2 | Main experiments

Model complexity and accuracy

When comparing different families of models, it is crucial to assess both their accuracy (measured by loss) and complexity (numbers of trainable parameters).[53] To facilitate a fair comparison among the MLP, KAN, and LKAN models, we adjusted the network width (numbers of neurons) and network depth (numbers of hidden layers) to create networks with varying number of trainable parameters. Specifically, by varying the network width, we categorized the models into six groups—small, small+, medium, medium+, large, and large+. Within each category, we trained and compared one-layer MLP, KAN, and LKAN models that had very similar numbers of trainable parameters. In addition, we trained MLP, KAN, and LKAN models with varying numbers of hidden layers, ranging from two to five, with 100 neurons in each hidden layer. The numbers of trainable parameters of these models were recorded. The results comparing the accuracy (validation loss) and complexity (numbers of trainable parameters) of the MLP, KAN, and LKAN models are presented in Section 3.2.1.

Cross validation, correlation analysis and reconstruction evaluation

To mitigate the effects of randomness that could arise from different data splits and to assess the stability of three

models, we used 5-fold cross-validation on MLP, KAN, and LKAN models (shape [43, 100, 9]).

To further evaluate the performance of MLP/KAN/LKAN, the independent test dataset including 51 977 Z-spectra from 2 participants was introduced. We applied the trained network models (shape [43, 100, 9]) with the lowest validation loss from five repeated training runs to extract ΔB_0 and multiple CEST fitting parameters from the unseen testing dataset and compared them with conventional MPLF. Voxel-by-voxel correlation analyses were conducted between the fitting results of MLP/KAN/LKAN models and the MPLF results.

Additionally, CEST MRI offers the advantage of producing CEST maps that contain spatial information related to molecules. Therefore, we used the voxel-wise CEST fitting parameters generated by the three network models to reconstruct CEST maps, and compared them with the MPLF results.

Furthermore, because the CEST fitting parameters are predicted from Z-spectra, it is also crucial to evaluate the ability of these three models to reversely reconstruct the Z-spectra from their generated CEST fitting parameters. To this end, we selected three regions of interest (ROIs) from different brain regions, including white matter, gray matter, and cerebrospinal fluid and then reconstructed the Z-spectra using the nine CEST fitting parameters generated by all three methods. The results of MLP/KAN/LKAN were compared to MPLF, and the absolute difference was calculated.

Results of the cross validation and correlation analysis experiments are displayed in Section 3.2.2.

2.3.3 | Noise performance evaluation

To compare the performance of different networks on noisy CEST Z-spectra data, we applied varying levels of noise to the train, validation, and test datasets. Given that input data was normalized to the range [0,1], we added Gaussian noise with a mean of 0 and a SD of 0.01, 0.02, 0.05, 0.1, and 0.2. A neural network with one hidden layer containing 100 neurons (shape [43, 100, 9]) was applied to all models. The comparison results for each network under varying noise levels are presented in Section 3.3.

2.3.4 | Interpretability investigations

Kernel Shapley additive explanations (SHAP)[59] is a widely adopted technique based on Shapley values derived from cooperative game theory. It effectively approximates the contribution of each input feature to the model's prediction. The primary advantages of Kernel SHAP include its model-agnostic nature, enabling transparent explanations of deep learning models. To rank the importance of individual input features on the MLP/KAN/LKAN model's predictions, we randomly selected 400 Z-spectra from test datasets as samples and applied the Kernel SHAP method to evaluate the contribution of 43 input features to 9 output variables. For a clearer understanding of feature impact, we created the global feature importance plots to visualize the top 20 most influential input features to the predicted ΔB_0, water/CEST amplitude values based on their mean absolute SHAP value. These visualization plots provided an intuitive representation of the most critical Z-spectral input features.

Furthermore, we successfully visualized the learnable activation functions of the KAN and LKAN models, which have a shape of [43, 4, 9] for demonstration purposes. This visualization of the learned univariate functions is expected to enhance the interpretability of both KAN and LKAN models.

The detailed results of interpretability investigations are displayed in Section 3.4.

3 | RESULTS

3.1 | Network hyperparameters

3.1.1 | GS for KAN

The specific results regarding the average validation loss for the KAN models with varying GS, along with the numbers of trainable parameters for each network are reported in Table 1A. The validation loss initially decreased from 0.42 and then stabilized around 0.38 as the GS increased from 1 to 10. With the increase in GS, the models were able to capture more detailed features, resulting in an improvement in overall loss. Notably, when GS equaled 9, the KAN model achieved the lowest validation loss in this study.

3.1.2 | GS and P for LKAN

The specific results regarding the average validation loss for LKAN with different GS and P, along with the numbers of trainable parameters for each model are presented in Table 1B. Based on the results, we proposed that LKAN yields superior results with GS = 1 and $P = 5$, where validation loss was the lowest at 0.3571.

3.2 | Main results

3.2.1 | Model complexity and accuracy

First, we compared the performance of models with different neuron size. According to the results presented in

TABLE 1 Records of parameter counts (the numbers of trainable parameters) and average validation loss from five repeated trainings with different hyperparameters in (A) KAN (grid size) and (B) LKAN (grid size and padding factor).

(A) CEST-KAN					
	GS = 1	**GS = 2**	**GS = 3**	**GS = 4**	**GS = 5**
Average validation loss	0.4245 ± 0.0038	0.4053 ± 0.0040	0.3971 ± 0.0036	0.3920 ± 0.0048	0.3857 ± 0.0045
Parameter count	31 200	36 400	41 600	46 800	52 000
	GS = 6	**GS = 7**	**GS = 8**	**GS = 9**	**GS = 10**
Average validation loss	0.3840 ± 0.0072	0.3828 ± 0.0036	0.3797 ± 0.0059	0.3780 ± 0.0057	0.3830 ± 0.0032
Parameter count	57 200	62 400	67 600	72 800	78 000

(B) CEST-LKAN						
		***P* = 1**	***P* = 3**	***P* = 5**	***P* = 7**	***P* = 9**
GS = 1	Average validation loss	0.3715 ± 0.0044	0.3659 ± 0.0056	0.3571 ± 0.0020	0.3641 ± 0.0084	0.3602 ± 0.0051
	Parameter count	26 395	47 195	67 995	88 795	109 595
GS = 3	Average validation loss	0.3909 ± 0.0082	0.3787 ± 0.0104	0.3641 ± 0.0027	0.3660 ± 0.0035	0.3667 ± 0.0072
	Parameter Count	57 595	119 995	182 395	244 795	307 195
GS = 5	Average validation loss	0.3984 ± 0.0090	0.3816 ± 0.0086	0.3804 ± 0.0090	0.3786 ± 0.0024	0.3784 ± 0.0104
	Parameter count	88 795	192 795	296 795	400 795	504 795

Abbreviations: GS, grid size; *P*, padding factor; KAN, Kolmogorov-Arnold network; LKAN, Lorentzian-KAN.

Table 2 and Figure 2A, the accuracy of the MLP, KAN, and LKAN models improved initially and then tended to stabilize as their complexity (number of trainable parameters) increased. When considering models with similar levels of complexity, the LKAN model consistently demonstrated the highest accuracy (lowest validation loss), closely followed by the KAN model, whereas the MLP consistently exhibited the lowest accuracy (highest validation loss). Notably, the MLP model with 100 709 trainable parameters (large+) had an average validation loss of 0.5954, whereas the KAN model with 2912 trainable parameters (small) could achieve an average validation loss of 0.4652, and the LKAN model with 2811 trainable parameters (small) could achieve an average validation loss of 0.4417. The comparison results highlight that the KAN and LKAN models can achieve higher accuracy with smaller model sizes. Additionally, as shown in Table 2 and Figure 2B,C, the training time per epoch and total training time of the MLP, KAN, and LKAN models did not change significantly with an increase in the numbers of neurons. The MLP models consistently exhibited the shortest training times, followed by the LKAN models. The KAN models, on the other hand, took roughly twice as long to train compared to the MLP models.

Next, we assessed the performance of models with varying numbers of hidden layers, each comprising 100 neurons. According to the results shown in Table 3 and Figure 2D, the accuracy for the MLP, KAN, and LKAN models exhibited similar trends. Specifically, there is a consistent pattern where accuracy initially increased before decreasing again as the hidden layer number increased. This pattern could be attributed to the networks gradually moving toward overfitting at higher complexities, which results in a decreased accuracy (an increased validation loss). As an illustration, the loss curves for both training and validation of MLP/KAN/LKAN models are displayed in Figures S1 and S2. When the number of hidden layers was set to one, all networks demonstrated rapid convergence, with loss curves that initially decreased quickly before plateauing (Figure S1). However, when the numbers of hidden layers of KAN increased to four, with a trainable parameter of 492 800 (Table 3), the KAN models have exhibited a possibility of overfitting (Figure S2). The KAN model's tendency to overfit more quickly than MLP can be explained by its faster neural scaling laws, which allow smaller models to achieve optimal performance more rapidly. Additionally, LKAN consistently exhibited the highest accuracy (the lowest validation loss) across all levels of complexity, whereas MLP showed the lowest accuracy. The validation loss for MLP was more sensitive to network depth compared to KAN and LKAN, the latter two demonstrating only slight changes across different numbers of layers before overfitting occurred. Despite this, the MLP model consistently had the shortest training time and was less affected by the increase in the number of hidden layers from 1 to 5 (Figure 2E). For models with a shape

TABLE 2 Training and validation results of MLP, KAN, and LKAN models with parameter counts (the numbers of trainable parameters) from five repeated trainings.

Neuron Size	Model	Parameter counts	Network shape	Average validation loss	Average training time (s)
Small	MLP	2924	[43, 55, 9]	0.6146 ± 0.0090	1039.2
	KAN	2912	[43, 4, 9]	0.4652 ± 0.0122	2028.0
	LKAN	2811	[43, 4, 9]	0.4417 ± 0.0134	1570.3
Small+	MLP	5309	[43, 100, 9]	0.5845 ± 0.0118	1134.8
	KAN	5824	[43, 8, 9]	0.4451 ± 0.0051	2028.6
	LKAN	5527	[43, 8, 9]	0.4019 ± 0.0164	1554.2
Medium	MLP	25 449	[43, 480,9]	0.5974 ± 0.0211	1104.4
	KAN	25 480	[43, 35, 9]	0.3924 ± 0.0037	2259.2
	LKAN	25 218	[43, 37, 9]	0.3620 ± 0.0033	1416.8
Medium+	MLP	68 008	[43, 1283, 9]	0.5827 ± 0.0133	1172.2
	KAN	67 704	[43, 93, 9]	0.3838 ± 0.0021	2227.4
	LKAN	67 995	[43, 100, 9]	0.3574 ± 0.0059	1591.6
Large	MLP	72 778	[43, 1373, 9]	0.5869 ± 0.0227	1159.8
	KAN	72 800	[43, 100, 9]	0.3796 ± 0.0040	2536.8
	LKAN	62 748	[43, 107, 9]	0.3583 ± 0.0047	1582.6
Large+	MLP	100 709	[43, 1900, 9]	0.5954 ± 0.0217	1168.6
	KAN	100 464	[43, 138, 9]	0.3789 ± 0.0015	2215.2
	LKAN	100 587	[43, 148, 9]	0.3584 ± 0.0086	1648.0

Abbreviations: KAN, Kolmogorov-Arnold network; LKAN, Lorentzian-KAN; MLP, multi-layer perception.

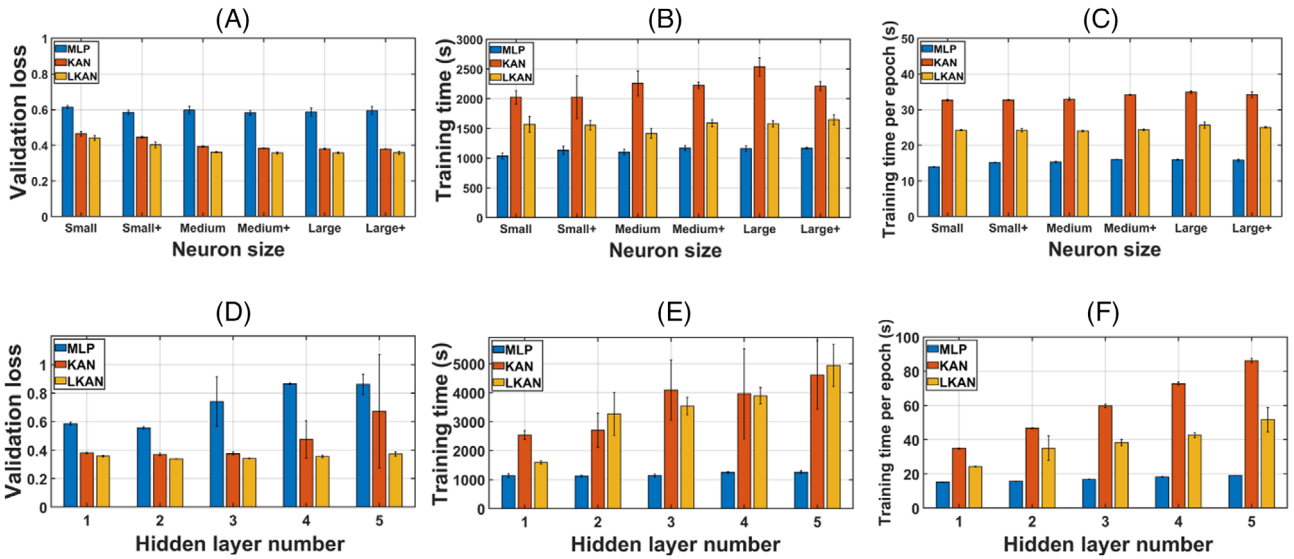

FIGURE 2 Average training and validation results of multi-layer perceptions (MLPs), Kolmogorov-Arnold network (KANs), and Lorentzian-KAN (LKANs) with different neurons size (A–C) and different numbers of hidden layers (D–F). The average results from five trials for each network are presented, with error bars indicating SD. (A) and (D) Validation loss. (B) and (E) Training time. (C) and (F) Training time per epoch.

TABLE 3 Training and validation results of from MLP, KAN, and LKAN models with varying number of hidden layers from five repeated trainings.

No. of hidden layers	Network architecture	Model	Parameter counts	Average validation loss	Average training time (s)
1	[43, 100, 9]	MLP	5309	0.5845 ± 0.0118	1134.8
		KAN	72 800	0.3796 ± 0.0040	2536.8
		LKAN	67 995	0.3574 ± 0.0059	1591.6
2	[43, 100, 100, 9]	MLP	15 409	0.5566 ± 0.0079	1135.9
		KAN	212 800	0.3694 ± 0.0095	2837.2
		LKAN	198 295	0.3380 ± 0.0014	3267.0
3	[43, 100, 100, 100, 9]	MLP	25 509	0.7416 ± 0.1732	1146.8
		KAN	352 800	0.3748 ± 0.0110	4090.2
		LKAN	328 595	0.3408 ± 0.0021	3536.8
4	[43, 100, 100, 100, 100, 9]	MLP	35 609	0.8659 ± 0.0063	1250.6
		KAN	492 800	0.4731 ± 0.1307	3962.2
		LKAN	458 898	0.3547 ± 0.0091	3891.8
5	[43, 100, 100, 100, 100, 100, 9]	MLP	45 709	0.8617 ± 0.0711	1254.2
		KAN	632 800	0.6722 ± 0.3980	4610.6
		LKAN	589 195	0.3720 ± 0.0158	4942.8

Abbreviations: KAN, Kolmogorov-Arnold network; LKAN, Lorentzian-KAN; MLP, multi-layer perception.

of [43, 100, 9], the training time for MLP was shorter than that for KAN and LKAN by 55.27% and 28.70%, respectively, and the LKAN models exhibited a shorter average training time than KAN by 37.26% (1134.8 vs. 2536.8 vs. 1591.6 s). The training time per epoch for LKAN was consistently shorter than that of KAN across varying numbers of hidden layers (Figure 2F). However, when the number of hidden layers was set to 2 and 5, the training time for LKAN exceeded that of KAN because of greater numbers of training epochs. As the network depth increased, LKAN was unable to maintain its significant advantage over KAN in terms of training time.

3.2.2 | Cross validation, correlation analysis and reconstruction evaluation

The comparison results of model performance across five folds for MLP, KAN, and LKAN are presented in Figure S3. KAN and LKAN consistently show lower validation loss compared to MLP. Additionally, the SD of average validation loss across the five folds for KAN and LKAN were 79.66% and 66.10% smaller than that of MLP (SD = 0.0012 vs. 0.0020 vs. 0.0059, respectively). Although KAN and LKAN exhibited lower SD compared to MLP, indicating superior stability, the overall validation losses were stable across the five folds for all models. Building on these results, we further evaluated the performance of the MLP, KAN, and LKAN models using the same initial k-fold settings.

Figure 3 displays the voxel-by-voxel correlation results of two testing subjects between the MLP/KAN/LKAN model and MPLF for ΔB_0 and other four CEST fitting parameters commonly used to generate water/CEST amplitude maps. The higher Pearson coefficients (R) in CEST fitting parameters supported that KAN and LKAN had superior performance over MLP in predicting CEST amplitude values (R = 0.9920 vs. 0.9943 vs. 0.9920, 0.9501 vs. 0.9905 vs. 0.9111, 0.9734 vs. 0.9901 vs. 0.9394, and 0.9976 vs. 0.9975 vs. 0.9944 for A_{water}, A_{amide}, A_{rNOE}, and A_{MT}, respectively). When considering five mentioned fitting parameters together, KAN and LKAN exhibited higher average Pearson coefficients than MLP by 1.57% and 2.84%, respectively, indicating superior performance. Simultaneously, when evaluated with the test dataset, both KAN and LKAN networks exhibited superior accuracy in predicting CEST parameters compared to the MLP. Specifically, KAN and LKAN achieved average reductions in test loss of 28.37% and 32.17%, respectively. The average test losses were recorded at 0.3006 for MLP, 0.2153 for KAN, and 0.2039 for LKAN.

Figure 4 displays the ΔB_0/water/CEST amplitude maps (ΔB_0, A_{water}, A_{amide}, A_{rNOE}, and A_{MT}) of testing subject 1 generated by all three methods. MLP, KAN, and LKAN

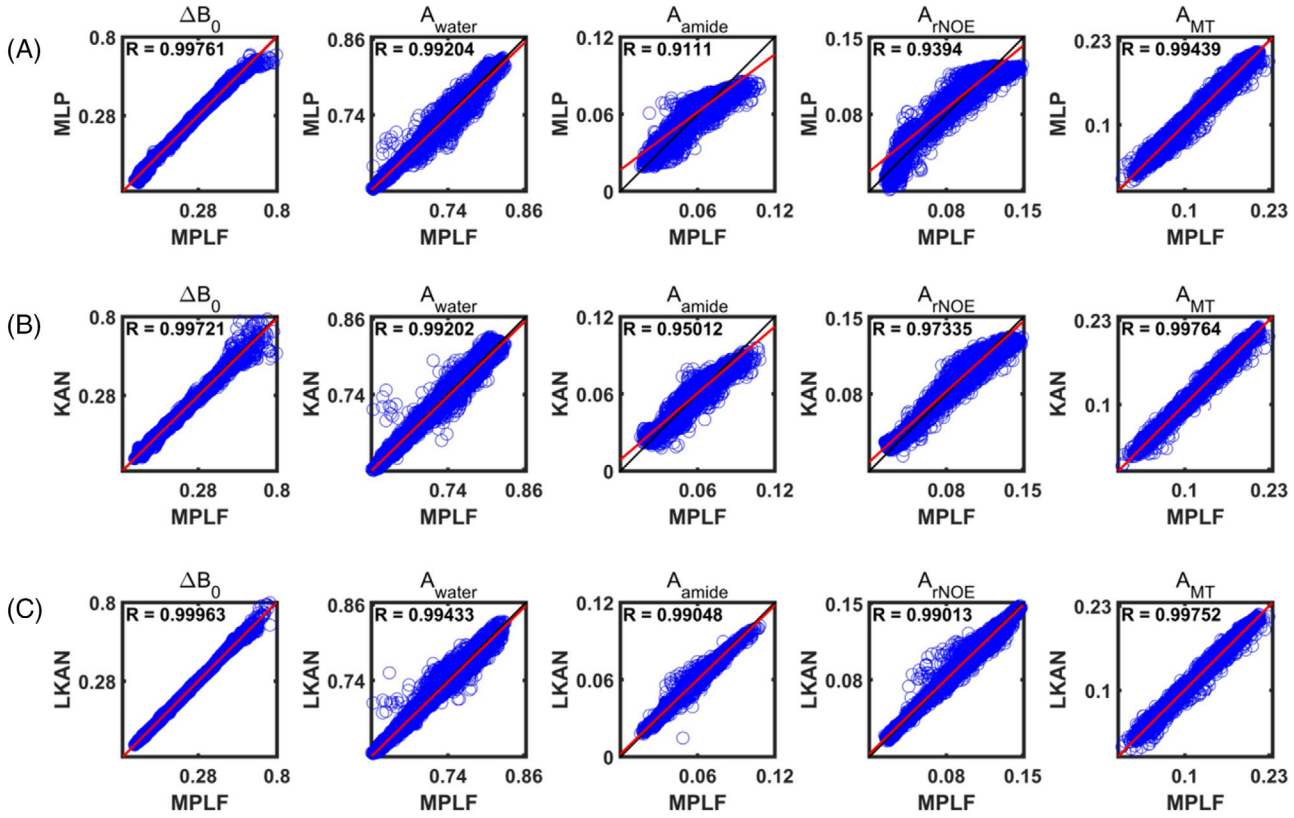

FIGURE 3 Application of the trained multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN) to test CEST dataset acquired from two human subjects. Correlation plots of predicted ΔB_0 (ppm) and amplitude values (a.u.) for (A) MLP, (B) KAN, and (C) LKAN versus multi-pool Lorentzian fitting (MPLF) fitting results (ground truth). R values in plots denote the Pearson correlation coefficient between prediction and ground truth. The columns represent the corresponding ΔB_0, water, amide, relayed nuclear Overhauser effect (rNOE), and magnetization transfer (MT) plots.

produced ΔB_0/water/CEST maps that were visually comparable to the results obtained using MPLF (Figure 4A–D). However, KAN and LKAN demonstrated higher accuracy than MLP, as evidenced by smaller errors displayed in the absolute difference maps (Figure 4E–G). On closer examination, we observed that the MLP results had a smoothing effect that removed small details in the maps, particularly for A_{amide}, A_{rNOE}, and A_{MT}. In contrast, the KAN and LKAN results retained these details, providing more information for the CEST maps. These findings were consistently observed in the results of testing subject 2 (Figure S4).

Figure 5 displays the Z-spectra of subject 1 reconstructed using the nine CEST fitting parameters generated by all three methods. The results of MLP/KAN/LKAN were compared to MPLF, and the absolute difference was calculated. In general, CSF exhibited the sharpest Z-spectra when compared to white matter and gray matter, as fluid-like tissue usually has smaller MT and direct water saturation effects when compared to solid-like tissues. Meanwhile, white matter exhibited a larger MT effect than gray matter, as it is composed of bundles of axons coated with myelin. All three methods captured these apparent features well. However, the comparison revealed that LKAN exhibited highest accuracy than MLP and KAN in all three regions, as evidenced by the smaller absolute difference when using MPLF as a reference. These observations were also found in the results of subject 2 (Figure S5).

Overall, the above presented results proved that LKAN and KAN (shape [43, 100, 9]) models significantly outperformed the MLP (shape [43, 100, 9]) models in predicting CEST fitting parameters.

3.3 | Model performance on noisy data

Test loss of MLP, KAN and LKAN at the noise levels of 0.01, 0.02, 0.05, 0.1, and 0.2 was 0.4284 versus 0.3795 versus 0.3727; 0.4453 versus 0.4210 versus 0.4205; 0.4862 versus 0.4729 versus 0.4677; 0.5602 versus 0.5344 versus 0.5248; 0.6592 versus 0.6601 versus 0.6402, respectively. As the noise level increased from 0.01 to 0.2 the test losses of MLP, KAN, and LKAN increased by 53.87%, 73.94%,

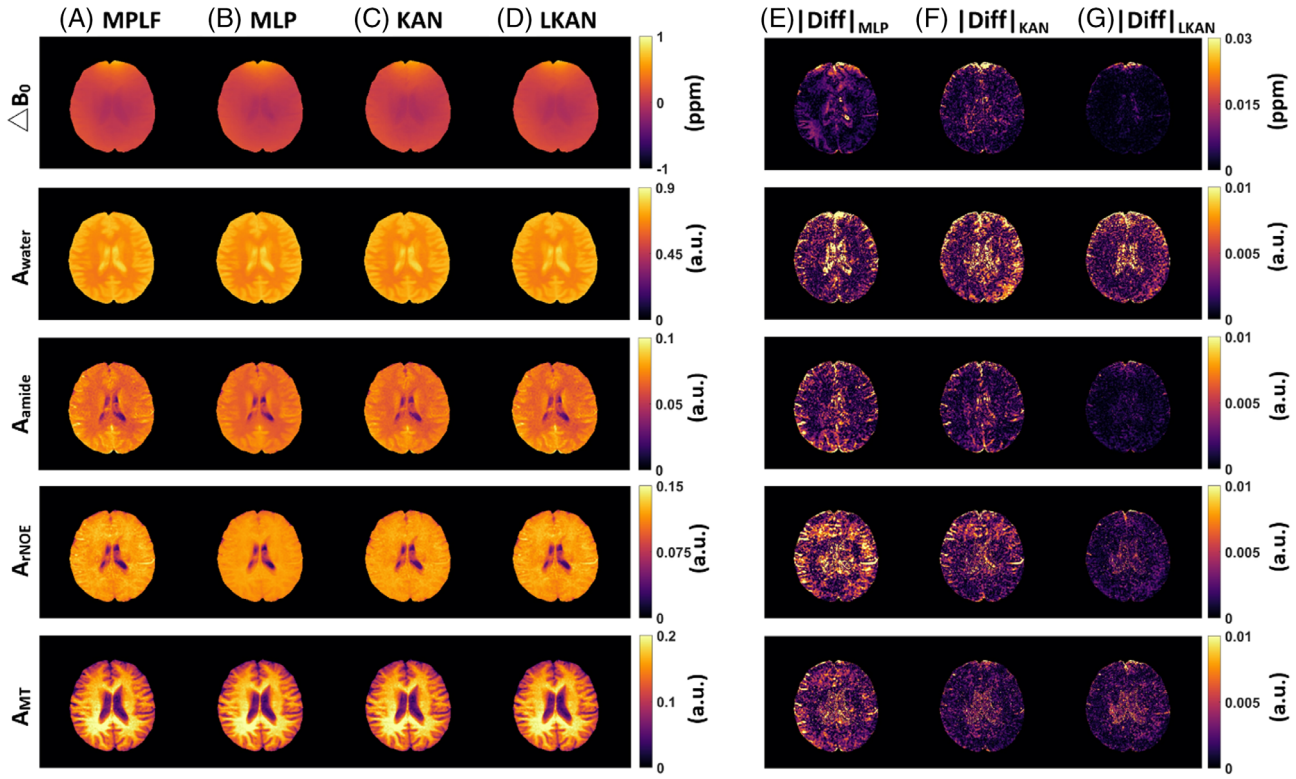

FIGURE 4 ΔB_0, Water and CEST amplitude maps of multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN) applied to the testing data of subject 1 that was not included in training, compared to multi-pool Lorentzian fitting (MPLF) fitting results. ΔB_0 was measured in ppm and amplitude values were in arbitrary units (a.u.). (A) ΔB_0, water and CEST amplitude maps obtained by MPLF method. (B) ΔB_0, Water and CEST amplitude maps obtained by MLP method. (C) ΔB_0, water and CEST amplitude maps obtained by KAN method. (D) ΔB_0, Water and CEST amplitude maps obtained by LKAN method. (E–G) Absolute difference (|Diff|) maps between the MPLF fitting outcomes and the predictions of the MLP, KAN, and LKAN models, respectively. The rows represent the corresponding ΔB_0, water, amide, relayed nuclear Overhauser effect (rNOE), and magnetization transfer (MT) amplitude maps.

and 71.77%, respectively. As shown in Figure 6F, both KAN and LKAN consistently outperformed MLP in accuracy for noise levels up to 0.1. However, at the highest noise level (0.2), where the data became highly distorted (Figure 6E), the performance gap between MLP, KAN, and LKAN narrowed, suggesting a common challenge for all three networks to capture key information. Overall, the results indicate that both KAN and LKAN demonstrate greater robustness to noisy data compared to MLP.

3.4 | Network interpretability

Figure 7 displays global feature importance plots, illustrating the top 20 input features that have the greatest contribution to the predicted values of water, ΔB_0, and CEST amplitudes. Notably, as shown in Figure 7C–E, KAN/LKAN models showed a significant increase in the mean absolute SHAP values in the prediction of the CEST amplitudes (A_{amide}, A_{rNOE}, and A_{MT}) compared to the MLP models, indicating a more effective utilization of input features. This enhanced contribution from input features may be a key reason why KAN/LKAN were more accurate (higher Pearson coefficients in Figure 3) in generating A_{amide}, A_{rNOE}, and A_{MT} and achieved a lower test loss than MLP. Furthermore, the MLP, KAN, and LKAN models all exhibited a higher utilization of inputs around 0 ppm when predicting ΔB_0 (Figure 7B), indicating that they effectively capture the DS effect, which aligns with the underlying CEST physics.

Figure S6 shows the learned univariate activation functions of the KAN and LKAN models. These visualizations illustrate how the networks transform input features through learnable activation functions, offering insights into the differences between KAN and LKAN in capturing complex characteristics. Specifically, a smaller number of activation functions actively contribute in LKAN, whereas the majority of activation functions in KAN have a more pronounced impact on the network. We speculate that this variance originates from the fact that the Lorentzian functions used in LKAN facilitate a more straightforward approximation of the Lorentzian curves in MPLF

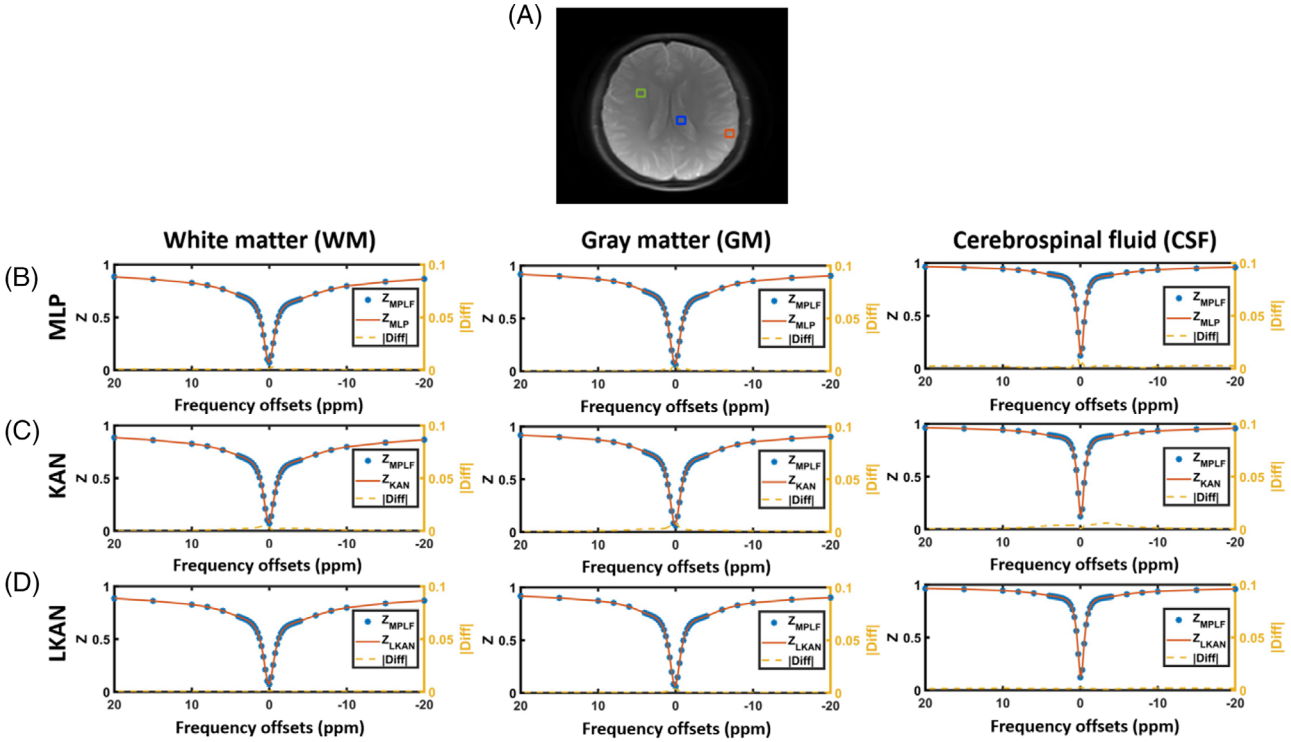

FIGURE 5 Comparison of the Z-spectra in the testing data of subject 1, reconstructed using the CEST fitting parameters predicted by multi-pool Lorentzian fitting (MPLF), multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN). (A) Three regions of interest (ROIs) used for generating the Z-spectra of white matter (green), gray matter (orange) and cerebrospinal fluid (blue). (B) Comparison of MLP with MPLF. (C) Comparison of KAN with MPLF. (D) Comparison of LKAN with MPLF.

of Z-spectra compared to spline functions, consequently leading to a decreased requirement for activation functions. This enhances the transparency of the networks and offers deeper insights into their underlying mechanisms, thereby improving our overall understanding of both KAN and LKAN models. Future in-depth studies in this perspective will be valuable for further enhancing the interpretation of the underlying mechanisms.

4 | DISCUSSION

In this study, we explored the potential of using KAN and the newly proposed LKAN to derive multiple CEST fitting parameters from Z-spectra, which have traditionally been obtained using the MPLF method or MLP models.

In comparison to the time-consuming MPLF, which took approximately 5 minutes to process one-slice brain CEST data, all these three network-based methods could quickly produce results in a second after training. The average inference time for two testing subjects of MLP, KAN, and LKAN was only 0.870, 0.893, and 0.865 seconds, respectively.

Furthermore, our results showed that KAN outperformed MLP in generating more accurate CEST fitting parameters from the same human CEST dataset. Specifically, KAN exhibited lower test losses and improved accuracy in predicting CEST fitting parameters and reconstructing CEST maps and Z-spectra (Figures 2–5).

This enhancement in accuracy could be attributed to innovative architectural design of KAN, which combines the strengths of spline functions and MLP networks. Specifically, the internal spline functions within KAN confer high accuracy for low-dimensional functions, while the external similarity to MLPs serves to alleviate the curse of dimensionality. Additionally, the activation functions of KAN are positioned at the edges rather than within the neurons, thereby enhancing its interpretability.[53]

Although KAN has its advantages, B-splines are not directly linked to CEST physics and incur high computational costs in the context of CEST Z-spectra analysis, which motivates us to replace B-splines with Lorentzian functions. Notably, LKAN consistently maintains lower test loss compared to KAN, indicating superior Z-spectrum fitting results in this study. The advantages of LKAN over KAN lie not only in test loss gains but also in the interpretability of the model.

By directly using Lorentzian functions, which are closely associated with the resonant peaks observed in Z-spectra, LKAN offers more meaningful insights into

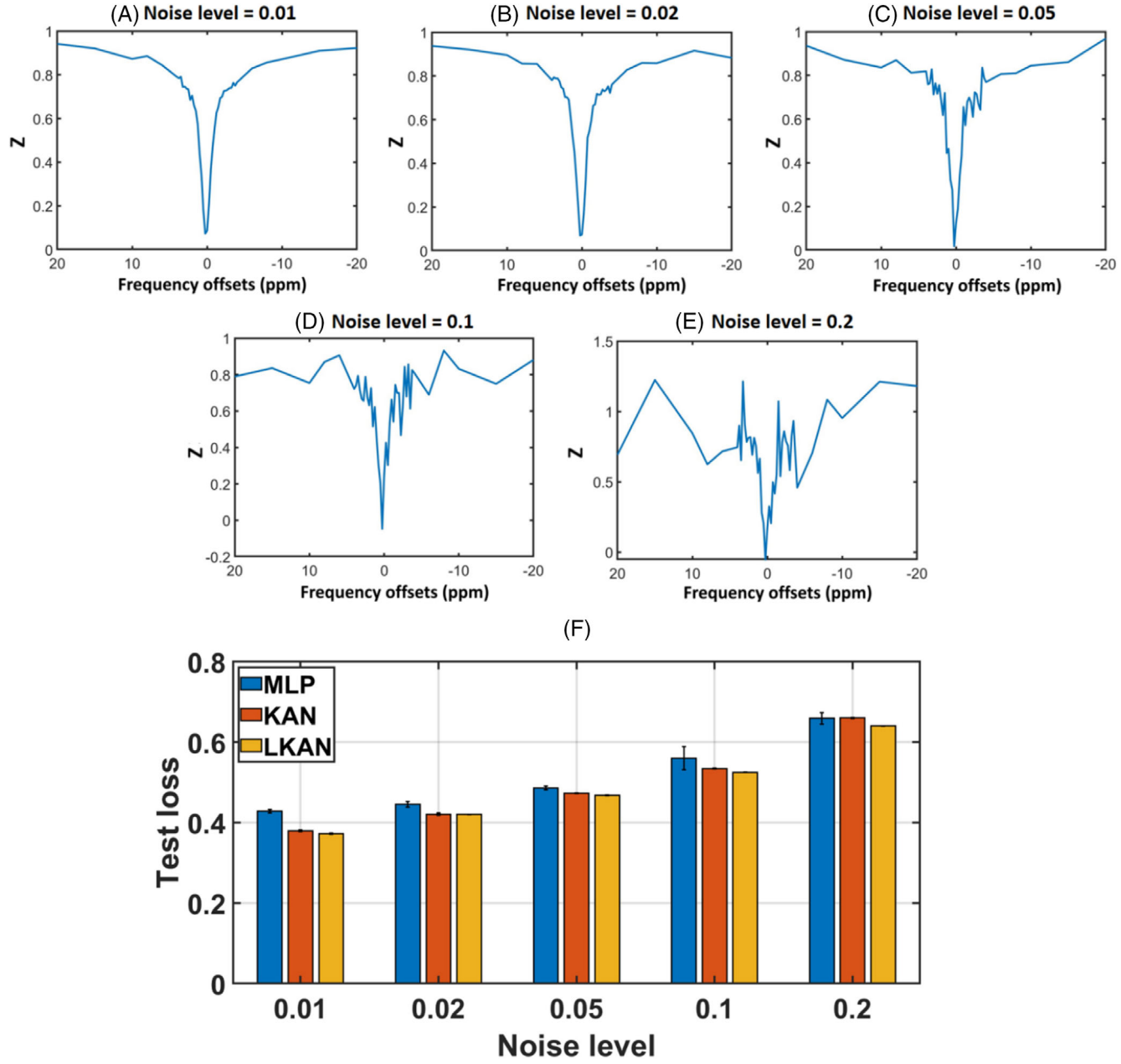

FIGURE 6 Average training results from five trials of multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN) at different noise levels, with error bars indicating SD. (A–E) Demonstrations of Z-spectra at the noise levels of 0.01, 0.02, 0.05, 0.1, and 0.2, respectively. (F) Test loss.

the data. For instance, by identifying the most critical Z-spectral features, we can focus on acquiring data at the most informative frequencies, thereby accelerating the CEST-MRI process without compromising diagnostic quality. Furthermore, LKAN achieves a 37.26% improvement in training efficiency over KAN by eliminating the need for computationally intensive B-spline functions, making it a more practical choice without sacrificing performance. Therefore, LKAN is not only a better fit, but also a more interpretable and practical tool for researchers and clinicians.

At the same time, the ability of KAN and LKAN to incorporate learnable activation functions provides a unique advantage, as it allows for direct visualization of how input features are dynamically transformed. This enhances the transparency of the networks and offers deeper insights into their underlying mechanisms, thereby improving our overall understanding of both KAN and LKAN models.

The training process for KAN/LKAN was considerably slower than that of MLP. However, as shown in this study, smaller KAN/LKAN models could achieve

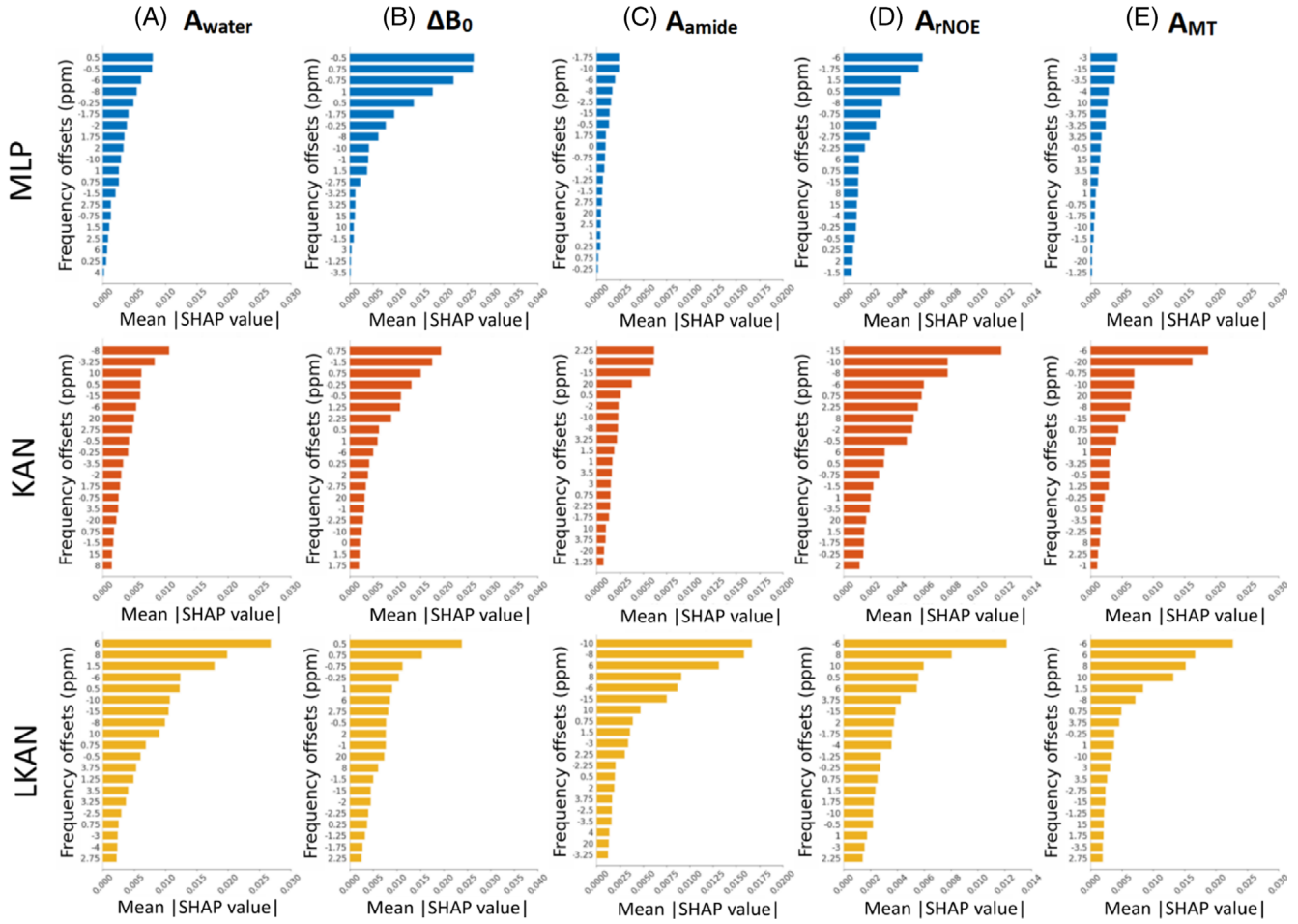

FIGURE 7 Global feature importance plot visualizing the top 20 most influential input features (based on the mean absolute Shapley additive explanations [SHAP] values) to the predicted outputs of (A) amplitude value of water, (B) ΔB_0 (ppm), (C) amplitude value of amide (a.u.), (D) amplitude value of relayed nuclear Overhauser effect (rNOE) (a.u.), and (E) amplitude value of magnetization transfer (MT) (a.u.). The rows represent the corresponding multi-layer perception (MLP), Kolmogorov-Arnold network (KAN), and Lorentzian-KAN (LKAN).

better accuracy than larger MLPs on the same tasks. This suggests that, despite the challenges related to training time, KAN/LKAN's efficient training and scaling properties could render it a viable alternative for a range of applications.

KAN and LKAN offer a highly promising approach for tackling the complex data analysis of CEST MRI. Currently, there is no standardized CEST MRI data analysis method. Although MPLF is a commonly used method, it cannot generate real ground truth for CEST MRI.[38,44] In this study, we used MPLF as an example to demonstrate the feasibility of using LKAN and KAN for CEST MRI data analysis.

We anticipate that KAN and LKAN can also be used in other CEST MRI techniques that use MLP for data analysis.[39–42] They hold potential for application in CEST MRF for extracting quantitative CEST fitting parameters, including concentration and exchange rate, from specially designed CEST-MRF spectra.[45–51] Furthermore, we expect that KAN and LKAN will have broad applicability in analyzing data from other MRI techniques that rely on MLP for postprocessing.[60–65]

5 | CONCLUSION

This study demonstrated the feasibility of using KAN and LKAN for CEST MRI data analysis in the human brain for the first time, with the KAN and LKAN model surpassing the conventional MLP approach in the accuracy of predicting the multiple CEST fitting parameters. Moreover, similar to other network-based methods, KAN and LKAN could generate CEST contrast maps at a much faster speed than the conventional MPLF method. These findings suggest that CEST-KAN and CEST-LKAN has the potential to be a robust and reliable postprocessing tool for CEST MRI in clinical settings, potentially overcoming the limitations of the current MPLF or MLP-based methods.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (82402225) and The University of Hong Kong (109000487, 109001694, 204610401, and 204610519).

DATA AVAILABILITY STATEMENT

The MPLF code for CEST MRI analysis is publicly available at: https://github.com/JianpanHuang/CEST-MPLF. The MLP, KAN, and LKAN codes for CEST MRI analysis are publicly available at: https://github.com/JianpanHuang/CEST-KAN.

ORCID

Huabin Zhang https://orcid.org/0000-0002-5170-2753
Jianpan Huang https://orcid.org/0000-0002-4453-8764

REFERENCES

1. Wolff SD, Balaban RS. Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo. *Magn Reson Med*. 1989;10:135-144.
2. Guivel-Scharen V, Sinnwell T, Wolff S, Balaban R. Detection of proton chemical exchange between metabolites and water in biological tissues. *J Magn Reson*. 1998;133:36-45. doi:10.1006/jmre.1998.1440
3. Ward K, Aletras A, Balaban RS. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J Magn Reson*. 2000;143:79-87. doi:10.1006/jmre.1999.1956
4. Van Zijl PC, Yadav NN. Chemical exchange saturation transfer (CEST): what is in a name and what isn't? *Magn Reson Med*. 2011;65:927-948. doi:10.1002/mrm.22761
5. Zhou J, Lal B, Wilson DA, Laterra J, van Zijl PC. Amide proton transfer (APT) contrast for imaging of brain tumors. *Magn Reson Med*. 2003;50:1120-1126. doi:10.1002/mrm.10651
6. Zhou J, Payen J-F, Wilson DA, Traystman RJ, van Zijl PC. Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nat Med*. 2003;9:1085-1090. doi:10.1038/nm907
7. Sui R, Chen L, Li Y, et al. Whole-brain amide CEST imaging at 3T with a steady-state radial MRI acquisition. *Magn Reson Med*. 2021;86:893-906.
8. Zaiss M, Windschuh J, Goerke S, et al. Downfield-NOE-suppressed amide-CEST-MRI at 7 tesla provides a unique contrast in human glioblastoma. *Magn Reson Med*. 2017;77:196-208.
9. Cai K, Haris M, Singh A, et al. Magnetic resonance imaging of glutamate. *Nat Med*. 2012;18:302-306.
10. Kogan F, Singh A, Debrosse C, et al. Imaging of glutamate in the spinal cord using GluCEST. *Neuroimage*. 2013;77:262-267.
11. Davis KA, Nanga RPR, Das S, et al. Glutamate imaging (GluCEST) lateralizes epileptic foci in nonlesional temporal lobe epilepsy. *Sci Transl Med*. 2015;7: 309ra161.
12. Bagga P, Crescenzi R, Krishnamoorthy G, et al. Mapping the alterations in glutamate with Glu CEST MRI in a mouse model of dopamine deficiency. *J Neurochem*. 2016;139:432-439. doi:10.1111/jnc.13771
13. Haris M, Nanga RPR, Singh A, et al. Exchange rates of creatine kinase metabolites: feasibility of imaging creatine by chemical exchange saturation transfer MRI. *NMR Biomed*. 2012;25:1305-1309.
14. Cai K, Singh A, Poptani H, et al. CEST signal at 2 ppm (CEST@2ppm) from Z-spectral fitting correlates with creatine distribution in brain tumor. *NMR Biomed*. 2015;28:1-8.
15. Chen L, Zeng H, Xu X, et al. Investigation of the contribution of total creatine to the CEST Z-spectrum of brain using a knockout mouse model. *NMR Biomed*. 2017;30:e3834.
16. Cai K, Tain R-W, Zhou XJ, et al. Creatine CEST MRI for differentiating gliomas with different degrees of aggressiveness. *Mol Imaging Biol*. 2017;19:225-232.
17. Chen L, Barker PB, Weiss RG, van Zijl PC, Xu J. Creatine and phosphocreatine mapping of mouse skeletal muscle by a polynomial and Lorentzian line-shape fitting CEST method. *Magn Reson Med*. 2019;81:69-78. doi:10.1002/mrm.27514
18. Chan K, McMahon M, Kato Y, et al. Natural D-glucose as a biodegradable MRI contrast agent for detecting cancer. *Magn Reson Med*. 2012;68:1764-1773. doi:10.1002/mrm.24520
19. Walker-Samuel S, Ramasawmy R, Torrealdea F, et al. In vivo imaging of glucose uptake and metabolism in tumors. *Nat Med*. 2013;19:1067-1072.
20. Xu X, Chan KW, Knutsson L, et al. Dynamic glucose enhanced (DGE) MRI for combined imaging of blood-brain barrier break down and increased blood volume in brain cancer. *Magn Reson Med*. 2015;74:1556-1563. doi:10.1002/mrm.25995
21. Huang J, Lai JH, Han X, et al. Sensitivity schemes for dynamic glucose-enhanced magnetic resonance imaging to detect glucose uptake and clearance in mouse brain at 3 T. *NMR Biomed*. 2022;35:e4640.
22. Huang J, van Zijl PC, Han X, et al. Altered d-glucose in brain parenchyma and cerebrospinal fluid of early Alzheimer's disease detected by dynamic glucose-enhanced MRI. *Sci Adv*. 2020;6:eaba3884. doi:10.1126/sciadv.aba3884
23. Chen L, Wei Z, Chan KW, et al. D-glucose uptake and clearance in the tauopathy Alzheimer's disease mouse brain detected by on-resonance variable delay multiple pulse MRI. *J Cereb Blood Flow Metab*. 2021;41:1013-1025. doi:10.1177/0271678X20941264
24. Jones CK, Huang A, Xu J, et al. Nuclear Overhauser enhancement (NOE) imaging in the human brain at 7 T. *Neuroimage*. 2013;77:114-124. doi:10.1016/j.neuroimage.2013.03.047
25. Desmond KL, Moosvi F, Stanisz GJ. Mapping of amide, amine, and aliphatic peaks in the CEST spectra of murine xenografts at 7 T. *Magn Reson Med*. 2014;71:1841-1853.
26. Xu X, Yadav NN, Zeng H, et al. Magnetization transfer contrast-suppressed imaging of amide proton transfer and relayed nuclear overhauser enhancement chemical exchange saturation transfer effects in the human brain at 7T. *Magn Reson Med*. 2016;75:88-96.
27. Goerke S, Soehngen Y, Deshmane A, et al. Relaxation-compensated APT and rNOE CEST-MRI of human brain tumors at 3 T. *Magn Reson Med*. 2019;82:622-632.
28. Huang J, Han X, Chen L, Xu X, Xu J, Chan KW. Relayed nuclear Overhauser enhancement imaging with magnetization transfer contrast suppression at 3 T. *Magn Reson Med*. 2021;85:254-267. doi:10.1002/mrm.28433

29. Huang J, Xu J, Lai JH, et al. Relayed nuclear Overhauser effect weighted (rNOEw) imaging identifies multiple sclerosis. *Neuroimage Clin*. 2021;32:102867. doi:10.1016/j.nicl.2021.102867
30. Zhou J, Zaiss M, Knutsson L, et al. Review and consensus recommendations on clinical APT-weighted imaging approaches at 3T: application to brain tumors. *Magn Reson Med*. 2022;88:546-574.
31. Jones CK, Polders D, Hua J, et al. In vivo three-dimensional whole-brain pulsed steady-state chemical exchange saturation transfer at 7 T. *Magn Reson Med*. 2012;67:1579-1589.
32. Zaiß M, Schmitt B, Bachert P. Quantitative separation of CEST effect from magnetization transfer and spillover effects by Lorentzian-line-fit analysis of z-spectra. *J Magn Reson*. 2011;211:149-155.
33. Zhang XY, Wang F, Li H, et al. Accuracy in the quantification of chemical exchange saturation transfer (CEST) and relayed nuclear Overhauser enhancement (rNOE) saturation transfer effects. *NMR Biomed*. 2017;30:e3716.
34. Wang K, Huang J, Ju L, et al. Creatine mapping of the brain at 3T by CEST MRI. *Magn Reson Med*. 2023;91:51-60. doi:10.1002/mrm.29876
35. Jin T, Wang P, Zong X, Kim SG. MR imaging of the amide-proton transfer effect and the pH-insensitive nuclear Overhauser effect at 9.4 T. *Magn Reson Med*. 2013;69:760-770.
36. Heo HY, Zhang Y, Jiang S, Lee DH, Zhou J. Quantitative assessment of amide proton transfer (APT) and nuclear overhauser enhancement (NOE) imaging with extrapolated semisolid magnetization transfer reference (EMR) signals: II. Comparison of three EMR models and application to human brain glioma at 3 tesla. *Magn Reson Med*. 2016;75:1630-1639.
37. Heo HY, Zhang Y, Lee DH, Hong X, Zhou J. Quantitative assessment of amide proton transfer (APT) and nuclear Overhauser enhancement (NOE) imaging with extrapolated semi-solid magnetization transfer reference (EMR) signals: application to a rat glioma model at 4.7 tesla. *Magn Reson Med*. 2016;75:137-149.
38. Glang F, Deshmane A, Prokudin S, et al. DeepCEST 3T: robust MRI parameter determination and uncertainty quantification with neural networks—application to CEST imaging of the human brain at 3T. *Magn Reson Med*. 2020;84:450-466.
39. Chen L, Schär M, Chan KW, et al. In vivo imaging of phosphocreatine with artificial neural networks. *Nat Commun*. 2020;11:1072. doi:10.1038/s41467-020-14874-0
40. Zaiss M, Deshmane A, Schuppert M, et al. DeepCEST: 9.4 T chemical exchange saturation transfer MRI contrast predicted from 3 T data–a proof of concept study. *Magn Reson Med*. 2019;81:3901-3914.
41. Hunger L, Rajput JR, Klein K, et al. DeepCEST 7 T: fast and homogeneous mapping of 7 T CEST MRI parameters and their uncertainty quantification. *Magn Reson Med*. 2023;89:1543-1556.
42. Xiao G, Zhang X, Yang G, Jia Y, Yan G, Wu R. Deep learning to reconstruct quasi-steady-state chemical exchange saturation transfer from a non-steady-state experiment. *NMR Biomed*. 2023;36:e4940.
43. Viswanathan M, Yin L, Kurmi Y, Zu Z. Machine learning-based amide proton transfer imaging using partially synthetic training data. *Magn Reson Med*. 2024;91:1908-1922.
44. Huang J, Lai JH, Tse KH, et al. Deep neural network based CEST and AREX processing: application in imaging a model of Alzheimer's disease at 3 T. *Magn Reson Med*. 2022;87:1529-1545. doi:10.1002/mrm.29044
45. Perlman O, Farrar CT, Heo HY. MR fingerprinting for semisolid magnetization transfer and chemical exchange saturation transfer quantification. *NMR Biomed*. 2023;36:e4710.
46. Kang B, Kim B, Schär M, Park H, Heo HY. Unsupervised learning for magnetization transfer contrast MR fingerprinting: application to CEST and nuclear Overhauser enhancement imaging. *Magn Reson Med*. 2021;85:2040-2054.
47. Kim B, Schär M, Park H, Heo H-Y. A deep learning approach for magnetization transfer contrast MR fingerprinting and chemical exchange saturation transfer imaging. *Neuroimage*. 2020;221:117165.
48. Perlman O, Ito H, Herz K, et al. Quantitative imaging of apoptosis following oncolytic virotherapy by magnetic resonance fingerprinting aided by deep learning. *Nat Biomed Eng*. 2022;6:648-657.
49. Singh M, Jiang S, Li Y, Van Zijl P, Zhou J, Heo HY. Bloch simulator–driven deep recurrent neural network for magnetization transfer contrast MR fingerprinting and CEST imaging. *Magn Reson Med*. 2023;90:1518-1536.
50. Cohen O, Yu VY, Tringale KR, et al. CEST MR fingerprinting (CEST-MRF) for brain tumor quantification using EPI readout and deep learning reconstruction. *Magn Reson Med*. 2023;89:233-249.
51. Cohen O, Otazo R. Global deep learning optimization of chemical exchange saturation transfer magnetic resonance fingerprinting acquisition schedule. *NMR Biomed*. 2023;36:e4954.
52. Cunningham H, Ewart A, Riggs L, Huben R, Sharkey L. Sparse Autoencoders Find Highly Interpretable Features in Language Models. arXiv preprint arXiv:230908600 2023.
53. Liu Z, Wang Y, Vaidya S, et al. Kan: Kolmogorov-Arnold Networks. arXiv preprint arXiv:240419756 2024.
54. Tian Y, Zhang C, Guo Z, Zhang X, Chawla N. Learning mlps on graphs: a unified view of effectiveness, robustness, and efficiency. *Paper presented at: The Eleventh International Conference on Learning Representations*; 2022.
55. Haykin S. *Neural Networks: a Comprehensive Foundation*. Prentice Hall PTR; 1998.
56. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syste*. 1989;2:303-314.
57. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2:359-366.
58. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. arXiv preprint arXiv:171105101 2017.
59. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing System*. 2017;4768–4777.
60. Fang Z, Chen Y, Liu M, et al. Deep learning for fast and spatially constrained tissue quantification from highly accelerated data in magnetic resonance fingerprinting. *IEEE Trans Med Imaging*. 2019;38:2364-2374.
61. Feng L, Ma D, Liu F. Rapid MR relaxometry using deep learning: an overview of current techniques and emerging trends. *NMR Biomed*. 2022;35:e4416.
62. Kwon K, Kim D, Park H. A parallel MR imaging method using multilayer perceptron. *Med Phys*. 2017;44:6209-6224. doi:10.1002/mp.12600

63. Twellmann T, Lichte O, Nattkemper TW. An adaptive tissue characterization network for model-free visualization of dynamic contrast-enhanced magnetic resonance image data. *IEEE Trans Med Imaging*. 2005;24:1256-1266.
64. Lucht RE, Knopp MV, Brix G. Classification of signal-time curves from dynamic MR mammography by neural networks. *Magn Reson Imaging*. 2001;19:51-57. doi:10.1016/s0730-725x(01)00222-3
65. Luu HM, Kim DH, Kim JW, Choi SH, Park SH. qMTNet: accelerated quantitative magnetization transfer imaging with artificial neural networks. *Magn Reson Med*. 2021;85:298-308.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Training and validation loss curves of (A) MLP, (B) KAN and (C) LKAN with the number of hidden layer at one (100 neurons per layer) on the same CEST dataset.
Figure S2. Training and validation loss curves of (A) MLP, (B) KAN and (C) LKAN with the number of hidden layer at four (100 neurons per layer) on the same CEST dataset.
Figure S3. Validation loss of MLPs, KANs and LKANs across five folds of cross-validation. The average results from five trials for each network are presented, with error bars indicating standard deviation.
Figure S4. ΔB_0, Water and CEST amplitude maps of MLP, KAN, and LKAN applied to the testing data of subject #2 that was not included in training, compared to MPLF fitting results. ΔB_0 was measured in ppm and amplitude values were in arbitrary units (a.u.). (A) ΔB_0, water and CEST amplitude maps obtained by MPLF method. (B) ΔB_0, Water and CEST amplitude maps obtained by MLP method. (C) ΔB_0, water and CEST amplitude maps obtained by KAN method. (D) ΔB_0, Water and CEST amplitude maps obtained by LKAN method. (E)–(G) Absolute difference (|Diff|) maps between the MPLF fitting outcomes and the predictions of the MLP, KAN and LKAN models, respectively. The rows represent the corresponding ΔB_0, water, amide, rNOE, and MT amplitude maps.
Figure S5. Comparison of the Z-spectra in the testing data of subject 2, reconstructed using the CEST fitting parameters predicted by MPLF, MLP, KAN and LKAN. (A) Three ROIs used for generating the Z-spectra of white matter (green), gray matter (orange) and cerebrospinal fluid (blue). (B) Comparison of MLP with MPLF. (C) Comparison of KAN with MPLF. (D) Comparison of LKAN with MPLF.
Figure S6. Visualization of the activation functions connecting the input neurons to the hidden layer neurons in the (A) KAN and (B) LKAN models (shape [43, 4, 9]). Each small plot represents a learned univariate activation function connecting the input neurons to the hidden layer neuron. The columns correspond to the 43 input neurons, while the rows represent the 4 neurons in the hidden layer.
Table S1. Initial values and bound conditions for MPLF with pool number set to four.

How to cite this article: Wang J, Cai P, Wang Z, Zhang H, Huang J. CEST MRI data analysis using Kolmogorov-Arnold network (KAN) and Lorentzian-KAN (LKAN) models. *Magn Reson Med*. 2025;94:1301-1317. doi: 10.1002/mrm.30548