Title

Automated selection of myocardial inversion time with a convolutional neural network: Spatial temporal ensemble myocardium inversion network (STEMI-NET)

Journal

Magnetic Resonance in Medicine, 81(5)

ISSN

0740-3194

Authors

Bahrami, Naeim
Retson, Tara
Blansit, Kevin
et al.

Publication Date

2019-05-01

DOI

10.1002/mrm.27680

Peer reviewed

Received: 28 September 2018 | Revised: 9 January 2019 | Accepted: 9 January 2019

DOI: 10.1002/mrm.27680

RAPID COMMUNICATION

Magnetic Resonance in Medicine

Automated selection of myocardial inversion time with a convolutional neural network: Spatial temporal ensemble myocardium inversion network (STEMI-NET)

Naeim Bahrami[1,2,3] | **Tara Retson**[1] | **Kevin Blansit**[4] | **Kang Wang**[1] | **Albert Hsiao**[1,3,4]

[1]Department of Radiology, University of California, San Diego, California

[2]Department of Psychiatry, University of California, San Diego, California

[3]Center for Multimodal Imaging and Genetics, University of California, San Diego, California

[4]Department of Biomedical Informatics, University of California, San Diego, California

Correspondence
Naeim Bahrami, Center for Multimodal Imaging and Genetics, University of California, San Diego, CA 92122.
Emails: naeim.bahrami@gmail.com; nabahrami@ucsd.edu
Twitter: @Naeimbah

Funding information
GE Healthcare

Purpose: Delayed enhancement imaging is an essential component of cardiac MRI, which is used widely for the evaluation of myocardial scar and viability. The selection of an optimal inversion time (TI) or null point (TI_{NP}) to suppress the background myocardial signal is required. The purpose of this study was to assess the feasibility of automated selection of TI_{NP} using a convolutional neural network (CNN). We hypothesized that a CNN may use spatial and temporal imaging characteristics from an inversion-recovery scout to select TI_{NP}, without the aid of a human observer.

Methods: We retrospectively collected 425 clinically acquired cardiac MRI exams performed at 1.5 T that included inversion-recovery scout acquisitions. We developed a VGG19 classifier ensembled with long short-term memory to identify the TI_{NP}. We compared the performance of the ensemble CNN in predicting TI_{NP} against ground truth, using linear regression analysis. Ground truth was defined as the expert physician annotation of the optimal TI. In a backtrack approach, saliency maps were generated to interpret the classification outcome and to increase the model's transparency.

Results: Prediction of TI_{NP} from our ensemble VGG19 long short-term memory closely matched with expert annotation ($\rho = 0.88$). Ninety-four percent of the predicted TI_{NP} were within ± 36 ms, and 83% were at or after expert TI selection.

Conclusion: In this study, we show that a CNN is capable of automated prediction of myocardial TI from an inversion-recovery experiment. Merging the spatial and temporal characteristics of the VGG-19 and long short-term-memory CNN structures appears to be sufficient to predict myocardial TI from TI scout.

KEYWORDS
cardiac magnetic resonance imaging (cMRI), convolutional neural network (CNN), deep learning, inversion recovery time

1 | INTRODUCTION

Delayed enhancement imaging is an essential component of cardiac MRI and is used widely for the evaluation of myocardial scar and viability. The selection of an optimal inversion time (TI), known as the myocardial null point (TI_{NP}), to suppress the background myocardial signal is required to optimize image contrast in myocardial delayed enhancement (MDE) acquisitions.[1] Incorrect selection of TI_{NP} can impair diagnostic quality. In certain diffuse myocardial diseases such as amyloidosis, it may be difficult to identify a single optimal null point.[2,3] Furthermore, it is known that TI_{NP} varies after intravenous contrast administration, and is therefore time-sensitive.[1]

In practice, selection of myocardial TI is generally performed through visual inspection and selection of TI_{NP} from an inversion-recovery scout acquisition. This is dependent on the skill of a technologist or physician to select the optimal TI, which may not be readily available outside of specialized centers. Thus, several technical approaches have been proposed to address this issue. Gassenmaier et al investigated the feasibility of developing a T_1 mapping–based method for the selection of TI_{NP} for late-gadolinium enhancement cardiac MRI.[1] The phase-sensitive inversion-recovery technique has also been widely adopted to broaden the range of acceptable TIs for MDE and to avoid inversion artifact.[4,5] However, such methods still rely on visual inspection of an image series by a trained human observer to select an optimal myocardial TI.

Deep learning approaches, including convolutional neural networks (CNNs), have the potential to automate the selection of TI and are the current state-of-the-art technology for image classification,[6-11] segmentation,[12-17] localization,[18-22] and prediction.[23] The CNNs apply a variation of multilayer perceptron to identify the optimal representation from the input (raw) data without requiring prior feature selection.[24] These models are known for capturing and automatically weighing the most relevant features from images (e.g., edge, contrast, intensity, morphology) to perform a specific task. However, there is a dearth of temporal information in static CNN models. During the visual inspection of the inversion scout sequence, trained observers often sequentially view the series of images to identify TI_{NP}. Long short-term memory (LSTM) models are a building unit for layers of a recurrent neural network, which observe temporal features from multiple frames. These models have been applied successfully for the evaluation of dynamic temporal activities in object recognition.[25,26] Combining the spatial features from CNN and the temporal features from LSTM could therefore provide a comprehensive set of feature maps to optimally select the null point from a TI scout sequence.

The purpose of this study is to assess the feasibility of automated selection of TI_{NP} using a new deep learning approach combining the characteristics of a CNN and LSTM, using a model we refer to as STEMINet (SpatioTemporal Ensemble Myocardial Inversion Network). We hypothesized that our ensembled model may utilize spatial and temporal imaging characteristics from an inversion-recovery scout to select TI_{NP}, without the aid of a human observer.

TABLE 1 Patient demographics and reason of study

		Patients (n)
Sex	Male	268
	Female	157
Age range (years)	12-88 (average: 46)	
Indications	Myocardial scar/viability	119
	Amyloid	10
	Hypertrophic cardiomyopathy	30
	Other cardiomyopathy	100
	Mass	54
	Congenital heart disease	9
	Pericardial disease	13
	Myocarditis/Sarcoid	34
	Others	56

2 | METHODS

2.1 | Patient demographics and cardiac MRI

With HIPAA compliance and internal review board approval with institutional waiver of informed consent, we retrospectively collected 425 clinically acquired cardiac MRI exams performed at 1.5 T from 2012 to 2017 (age: 12-88 years, 157 female and 268 male) (Table 1), which included inversion-recovery scout acquisitions 8 minutes after the administration of gadolinium-based intravenous contrast. For all subjects, 0.3 mL/kg (0.3 mmol/kg) of gadobenate dimeglumine was administered. The inversion-recovery (cine IR) scout sequence captures image contrast evolution at multiple time points following an inversion pulse, and is typically used to identify the optimal TI for MDE imaging. The inversion-recovery scout contained 30 frames with flip angle = 10, matrix = 128 × 128, slice thickness = 8 mm, TR = 4.5 ms, and TE = 2.0 ms, acquired with a temporal resolution of 24 ms to 36 ms and TIs ranging from 88 ms to 1289 ms.

2.2 | Inversion time null point (TI_{NP}) ground-truth annotation

A radiology resident (T.R.) was trained to visually select the image in the sequence corresponding to the optimal myocardial TI (TI_{NP}) by a board-certified cardiac radiologist (A.H.) with 10 years of experience in cardiac MRI. Due

to the frequently suboptimal selection of TIs in our clinical exams, we chose the more reliable reference standard of a dedicated radiologist-supervised annotation of the cine IR images, rather than the selected TI of the MDE images. The optimal TI_{NP} was marked on all available inversion-recovery scout sequences as the time point at which the myocardium had the lowest homogeneous signal intensity, after resolution of inversion artifact.

2.3 | Problem definition and model design

Although it is tempting to define the problem of selecting a single optimal TI in a cine IR series as a classification task to identify a solitary frame from a series of time frames, we instead posed the problem as a more balanced 2-class classification problem: classifying frames before the optimal TI as "early," and those at or following the optimal TI as "acceptable." Image frames before the ground-truth TI_{NP} were considered "early," whereas the image frames at or beyond TI_{NP} were considered "acceptable." We believed that this more balanced 2-class formulation would be more tractable for a neural network. In addition, because we felt that the temporal context of each frame could be helpful for identifying the optimal TI, we implemented a sliding window of 4 consecutive frames as inputs to the neural network (Figure 1). Each window was assigned a class based on its third frame, so that 2 frames preceding and 1 frame following the frame of interest could provide context. In other words, the input data window spanned 4 image frames, including the frame of interest (t_0), 2 preceding (t_{-2}, t_{-1}), and one following (t_1). Formulating the problem in this way, predicted inversion time can be calculated as the third image of the first *acceptable* window in a time series.

2.4 | Model structure and training

To solve this classification task, we devised an ensemble neural network called the STEMI-Net (Figure 2). This network includes a deep VGG19 convolutional network,[27] with the primary intent of capturing relevant spatial features. In addition, we included a recurrent neural network in the form of LSTM[28] to address temporal features relevant to this task. We constructed and trained 3 neural networks separately to assess their relative performance (VGG19, LSTM, and STEMI-Net). All models were built based on the inversion-recovery scout images, and no transfer learning or weight initialization was used. The models were each independently trained on a GPU workstation running Ubuntu 16.04, equipped with a Titan X graphics card (NVIDIA, Mountain View, CA). For each model described subsequently, networks were used to classify windows into 2 classes, "early" and "acceptable," and the third frame defined the class.

2.4.1 | VGG19

First, we constructed a 4-channel VGG19 network, in which a sliding window of 4 consecutive frames was used as the

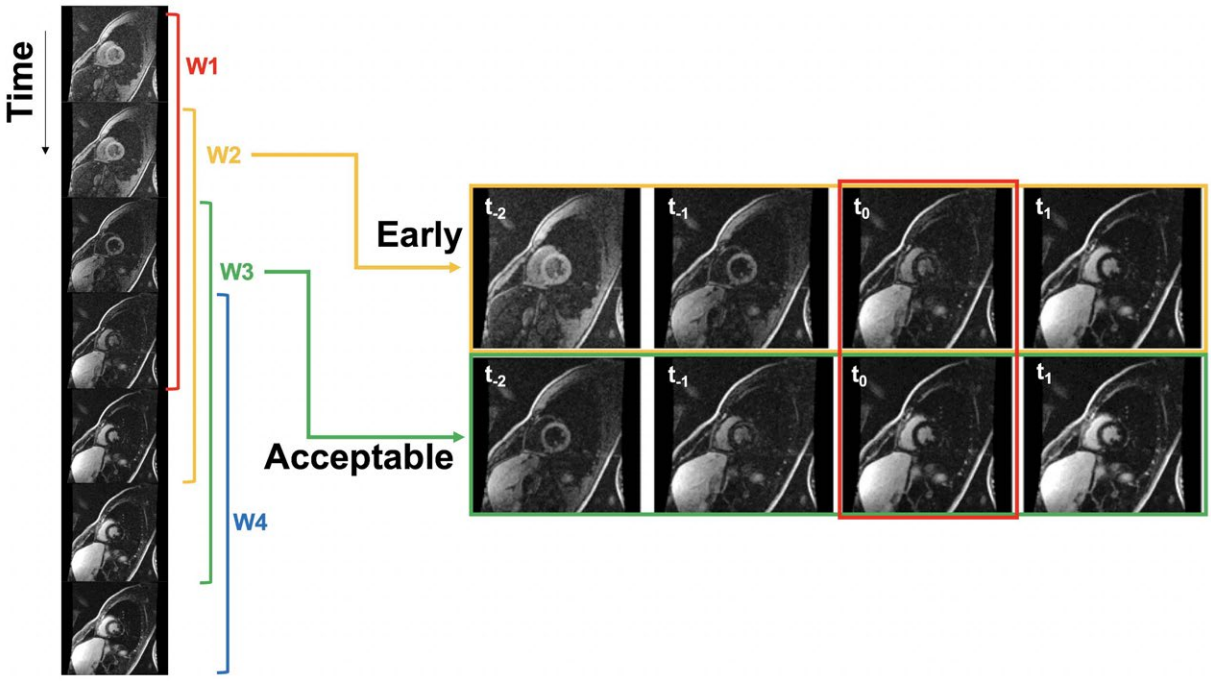

FIGURE 1 Formulation of the inversion time selection problem as a time window classification task. From a series of images from an inversion-recovery scout (left), multiple 4-frame image windows can be created. Two of these 4-frame image windows are shown (right). Frames preceding the optimal inversion time (TI) are labeled as "early," and frames at or following this time are labeled as "acceptable." On the right are 2 example windows. The third frame of each window (t_0) defines the class of the window

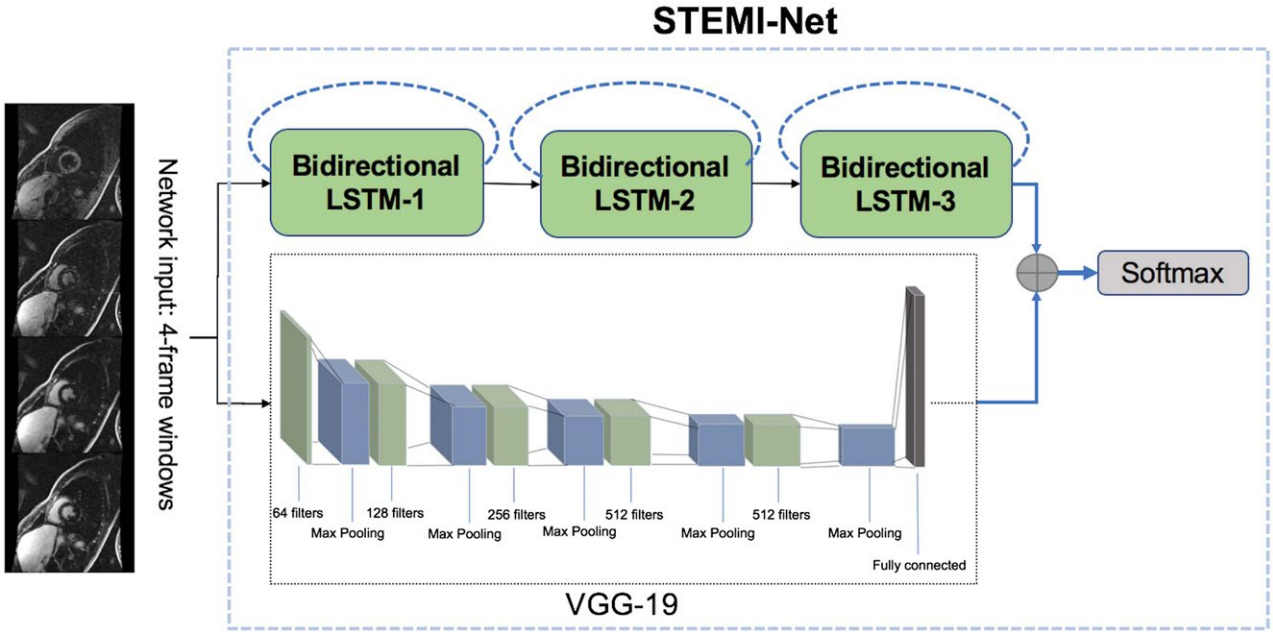

FIGURE 2 Structure of STEMI-Net (SpatioTemporal Ensemble Myocardial Inversion Network). The network architecture contains a VGG19 classifier (convolutional neural network) ensembled with long short-term memory (LSTM) block (recurrent neural network) to extract the most efficient and related spatial and temporal features on the cardiac MRI inversion-recovery scout. This model concatenates the features captured by convolutional neural network and recurrent neural network blocks to classify windows as early or acceptable, to identify the most optimal myocardial TI or null point

input. This model included 5 blocks of convolutional layers and was trained in the Keras environment with a Tensorflow backend. The initial hyperparameters of this network were set with a learning rate of 10^{-3}, momentum of 0.9, kernel size of 21 in the first block and 3 in the rest, decay of 10^{-4} of the learning rates, using stochastic gradient descent as the optimizer, while shuffling the data. These 5 blocks were able to create 1024 deep features for each individual input.

2.4.2 | Long short-term memory

We also constructed a separate recurrent neural network to address the temporal relationship between image frames using LSTM. This LSTM model included 16 filters, with a 2D kernel size of 3. We maintained the image size at 256 × 256 with the embedding dimensions of 128. Three serial bidirectional LSTMs (Keras LSTM model) with return sequence were implemented to capture both forward and backward relationships across the frames (classes were defined as "early" and "acceptable"; the third frame defined the class).

2.4.3 | STEMI-Net

Finally, we developed a third and final network, called STEMI-Net, in which the spatial features captured by a VGG19 were ensembled with the temporal features captured by 3 LSTM blocks. Each network extracted 1024 features independently, and concatenating these features provided us with a large number of parameters (n = 2048). Two dense layers were used to reduce the redundancies in the feature maps. We used 0.5 dropout to increase the robustness of the remaining features. Training error was minimized using the stochastic gradient descent optimizer, minimizing binary cross-entropy, and was validated by classification accuracy. Training was performed over 50 epochs with a batch size of 1.

2.5 | Saliency map

A saliency map, also known as an attention map, is a visual representation of salient regions, pixels, or objects in an image that are activated during the classification task.[29,30] Inspection of saliency maps can help localize the spatial and temporal characteristics that a CNN uses to make the final classification into early and acceptable categories. Saliency maps were created based on a backpropagation paradigm, decoding the most important features from the input image. The rectified linear unit was used as the backpropagation modifier for saliency map visualization, using the Keras_Vis toolkit.[31] Saliency maps were generated based on the temporal and spatial characteristics of all 4 slices of the windows to

illustrate relevant features, and then were overlaid on the t_0 slide for presentation.

2.6 | Statistics

Five-fold cross-validation was performed. For each cross-validation step, the neural networks were initialized with random weights (keras: kernel_initializer = "random_uniform"; bias_initializer = "zeros") and freshly and independently trained with 80% of the patients, whereas the remaining 20% of patients were used as independent test data. Cross-validation is a commonly used strategy to ensure that proposed algorithms are not dependent on the selection of any particular subset of training data. Network training performance was assessed by accuracy and model loss at each epoch. In the model prediction step, classification accuracy for each window was captured on the validation data set. The confusion matrix and its parameters were calculated on the validation data set in a per-patient scheme. We compared the performance of the ensemble CNN for predicting TI_{NP} against ground truth and calculated the correlation coefficient using the Pearson method. Ground truth was defined as the expert physician annotation of optimal TI. Python 3.6 and R were used for statistical analysis.

3 | RESULTS

We first evaluated the relative performance of the VGG19, LSTM, and STEMI-Net neural networks for classifying each of the 4-frame windows into early and acceptable categories. Using VGG19 alone, classification accuracy was 83.9% ± 2.8%. Using LSTM alone, classification accuracy was 81.9% ± 3.2%. In contrast, the ensemble STEMI-Net had a classification accuracy of 92.1% ± 2.2%, exceeding the performance of either of its component networks. STEMI-Net learned 419 870 850 parameters. Table 2 indicates the accuracy of all 3 models in 5-fold cross-validation.

TABLE 2 Training and accuracy of STEMI-NET and its component models

Model	Accuracy
VGG19	83.9% ± 2.8%
LSTM	81.9% ± 3.2%
STEMI-NET[a]	
Fold 1	92.9%
Fold 2	94.2%
Fold 3	93.9%
Fold 4	89.9%
Fold 5	89.8%
Total	92.1% ± 2.1%

[a]STEMI-Net outperformed the predictions of its individual components.

3.1 | Inversion-recovery null point prediction

Next, we evaluated the performance of the neural networks in identifying the myocardial null point. As described previously, the predicted null point can be calculated as the third frame (t_0) of the first acceptable window in a time series. Prediction of TI_{NP} from our ensemble STEMI-Net closely matched with expert annotation ($\rho = 0.88$). STEMI-Net predicted the exact inversion-recovery time as the ground truth for 63% of the patients (n = 285). In 94% of cases (n= 397), predictions of TI_{NP} were within 1 frame (approximately 36 ms) of the ground truth. In addition, in 83% of cases (n = 352) the prediction occurred at or after the ground truth, suggesting that the neural network was largely able to choose TIs at time points after inversion artifact. Figure 3 demonstrates the performance of STEMI-Net.

3.2 | Transparency and attention maps

To further explore the underlying characteristics of STEMI-Net, we created saliency maps to better understand

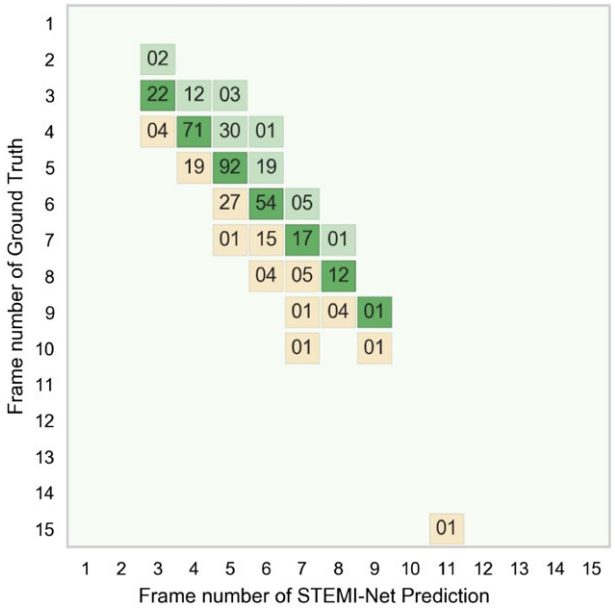

FIGURE 3 Performance of STEMI-Net relative to the ground truth. Dark green boxes show the number of patients for whom the optimal ground-truth frame exactly matched the STEMI-Net system. Sixty-three percent of predictions exactly matched ground truth and 94% of predictions were within 1 frame of the ground truth. The light green boxes show the number of patients in whom the prediction underestimated ground truth, and the yellow boxes show the number of patients in whom prediction overestimated ground truth

the behavior of the network during the classification task. These maps highlight the magnitude and location of the features with greatest activation during the classification task. Saliency map analysis on STEMI-Net revealed that the network attention focused primarily on myocardium and skeletal muscle on windows that it ultimately classified as early (Figure 4A). In contrast, attention appeared to be more diffuse for windows classified as acceptable, with slightly increased attention to the blood pool. In cases in which myocardial scar was present (Figure 4B), this

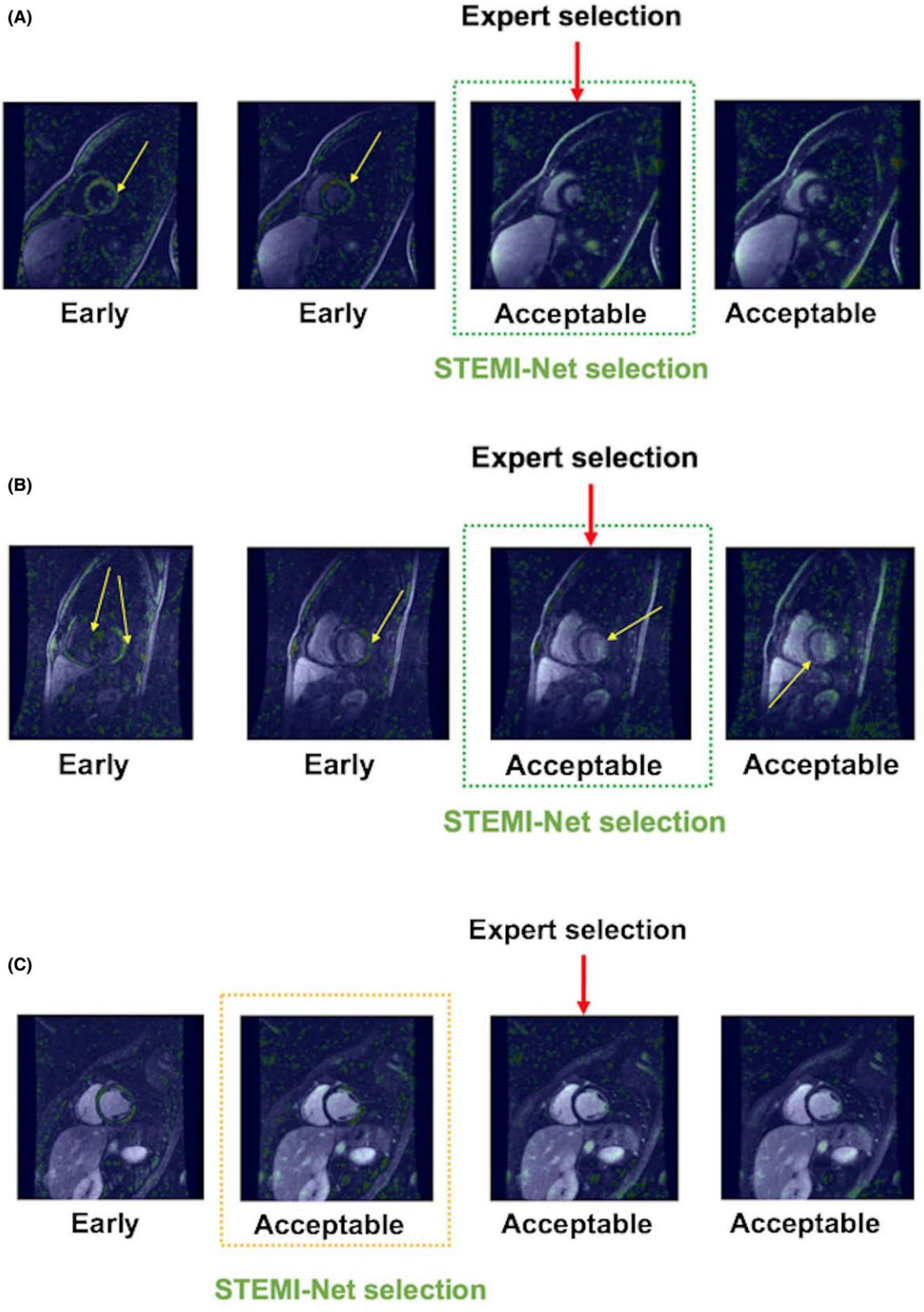

FIGURE 4 A, Saliency maps from a typical case, classified with STEMI-Net. On the top row are the source images from an inversion-recovery scout sequence. On the bottom row are saliency map color overlays on top of the source images, showing areas of highest attention by the neural network. Classification of windows as early is associated with attention to left ventricle and right ventricle myocardium, as well as skeletal muscle in the chest wall. Classification of windows as acceptable is associated with more diffuse saliency maps with relatively absent attention to the myocardium and slightly greater attention to the blood pool. B, Saliency maps from a case with myocardial delayed enhancement, classified with STEMI-Net. On the top row are the source images from an inversion-recovery scout sequence. On the bottom row are saliency map color overlays on top of the source images, showing areas of highest attention by the neural network. The neural network appears to ignore the area of myocardial scar in the septal wall when making the determination of an early time window. C, Saliency maps from a case with slight discordance between ground truth and STEMI-Net. On the top row are the source images from an inversion-recovery scout sequence. On the bottom row are saliency map color overlays on top of the source images, showing areas of highest attention by the neural network. STEMI-Net identified a time point 1 frame earlier than that annotated by an expert radiologist. In retrospect, this selection time point may be closer to the actual myocardial null point/TI

pattern persisted. The classification of early windows was performed with attention to segments of normal myocardium, and little or no attention to scar. In the few rare cases in which there was disagreement between STEMI-Net and the expert reader, occasionally this was due to a more accurate or equivocal prediction by the neural network (Figure 4C). In this example, STEMI-Net predicted the null point 1 frame earlier than the expert. In retrospect, both time points were comparable, and an expert reader could have just as readily chosen the earlier time frame.

4 | DISCUSSION

In this study, we showed that a CNN is capable of automated prediction of myocardial TI from an inversion-recovery experiment. We accomplish this through a novel transformation of an image-frame selection problem into a 2-class classification task. The 2-class formulation enables the use of a classification approach, in which an otherwise selection of a single, optimal time frame may be too imbalanced to solve with this strategy—analogous to searching for a "needle in a haystack." Merging the spatial and temporal characteristics of the VGG-19 and LSTM CNN structures appears to be sufficient to accurately classify these frames, and improves on each of the component networks. In this study, we studied the use of VGG19, which leverages spatial features, and LSTM, which leverages temporal features: a combination of both improved classification accuracy and overall null point prediction. Furthermore, study of the saliency maps of STEMI-Net confirms that the classification task is accomplished by monitoring the signal characteristics of the myocardium.

Study of the behavior of neural network through saliency maps may be valuable to assess their generalizability beyond the training data set. In this study, we calculated saliency maps through backpropagation, to localize features that contribute to the final prediction. Study of saliency maps showed that this model predicts the correct frame classification based on signal characteristics of myocardium, skeletal muscle, and blood pool. This network appears to be robust, even in the presence of scar in the myocardium, and suggests that it may have clinical value if incorporated into clinical MRI protocols.

To our knowledge, STEMI-Net is the first application of deep learning to address the selection of myocardial TI. A few deep learning algorithms have been proposed for other areas in cardiac MRI.[32] Previously, Avendi et al used the CNN to automatically detect the left ventricle chamber in an MRI data set.[32] In addition, Zhang et al combined recurrent neural network with convolutional LSTM for left-ventricle myocardium segmentation.[33] Xue et al introduced a spatial–temporal circle LSTM model to calculate left-ventricle myocardial thickness in the short axis scan.[34]

A few limitations of this feasibility study should be noted. In the training and validation of this model, we restricted training to short-axis inversion-recovery scout acquisitions. Because some sites prefer 4-chamber or other long-axis acquisitions for the inversion-recovery scout sequence, future work may incorporate long-axis inversion data. Similarly, for this initial study, we included only training data from 1.5T scanners from a single vendor at our institution, and incorporation of data acquired at higher field strength (e.g., 3 T) and across vendors may improve the generalizability of the model. We did not specifically address variations in radiologist preference for "optimal null point" in this work. It is possible that a different radiologist may systematically choose slightly earlier or later TIs, which raises the issue that an important design decision for neural networks can be the choice of ground truth. Because STEMI-Net may be readily retrained with an alternate set of ground-truth annotations, a future study could be performed to assess the use of ground truth obtained from multiple radiologists on the performance of the algorithm. Moreover, the inversion-recovery scout sequence has a finite temporal resolution, and it is possible that the optimal TI may lie between image frames of this sequence, which we did not fully explore here. Another limitation of the study is that there were relatively few cases of cardiac amyloidosis in our patient cohort, for which identification of a single TI may not be technically

feasible due to intrinsic tissue heterogeneity of the myocardium. Further investigation of how to optimize neural networks for this and other specific patient subgroups may be warranted. This might require obtaining an additional, larger cohort of such patients for further model training. Finally, a comprehensive test of the performance of the model in routine clinical practice would serve as the ultimate test of its applicability and effectiveness in addressing this long-standing problem.

5 | CONCLUSIONS

In this study, we showed that STEMI-Net is capable of automatically identifying the optimal myocardial TI from an inversion-recovery scout acquisition. Implementation of this ensemble CNN into a clinical workflow may facilitate the automation of TI and may help improve the reliability and accessibility of cardiac MRI.

ACKNOWLEDGMENTS

The authors acknowledge the support of NVIDIA Corporation for their generous donation of a Titan Xp GPU for training the convolutional neural networks in this research.

CONFLICTS OF INTEREST

A.H. is a founder, shareholder, and holder of intellectual property rights for Arterys. A.H. receives research grant support from GE Healthcare. A.H. is a speaker for and receives research grant support from Bayer. The current work was funded with research grant support from GE Healthcare.

ORCID

Naeim Bahrami https://orcid.org/0000-0001-7491-3302

REFERENCES

1. Gassenmaier S, van der Geest R, Schoepf UJ, et al. Quantitative inversion time prescription for myocardial late gadolinium enhancement using T1-mapping-based synthetic inversion recovery imaging: reducing subjectivity in the estimation of inversion time. *Int J Cardiovasc Imaging*. 2018;34:921–929.
2. Pandey T, Jambhekar K, Shaikh R, Lensing S, Viswamitra S. Utility of the inversion scout sequence (TI scout) in diagnosing myocardial amyloid infiltration. *Int J Cardiovasc Imaging*. 2013;29:103–112.
3. vanden Driesen RI, Slaughter RE, Strugnell WE. MR findings in cardiac amyloidosis. *AJR Am J Roentgenol*. 2006;186:1682–1685.
4. Huber AM, Schoenberg SO, Hayes C, et al. Phase-sensitive inversion-recovery MR imaging in the detection of myocardial infarction. *Radiology*. 2005;237:854–860.
5. Kellman P, Arai AE, McVeigh ER, Aletras AH. Phase-sensitive inversion recovery for detecting myocardial infarction using gadolinium-delayed hyperenhancement†. *Magn Reson Med*. 2002;47:372–383.
6. Margeta J, Criminisi A, Lozoya RC, Lee DC, Ayache N. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2017;5:339–349.
7. Zhang L, Gooya A, Dong B, et al. Automated quality assessment of cardiac MR images using convolutional neural networks. In: Tsaftaris SA, Gooya A, Frangi AF, Prince J, eds. *Simulation and synthesis in medical imaging*. Cham, Switzerland: Springer International Publishing; 2016:138–145.
8. Sarraf S, Tofighi G; Alzheimer's Disease Neuroimaging Initiative. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv*. 2016. doi: https://doi.org/10.1101/070441.
9. Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. In: Proceedings of the 14th International Symposium on Biomedical Imaging, Melbourne, Australia; 2017: 835–838.
10. Zhang R, Liu Q, Cui H, et al. Thyroid classification via new multi-channel feature association and learning from multi-modality MRI images. In Proceedings of the 15th International Symposium on Biomedical Imaging, Washington DC; 2018: 277–280.
11. Oksuz I, Ruijsink B, Puyol-Anton E, et al. 15 Automatic mis-triggering artefact detection for image quality assessment of cardiac MRI. *Heart*. 2018;104:A5.
12. Milletari F, Ahmadi S-A, Kroll C, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput Vis Image Underst*. 2017;164:92–102.
13. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61–78.
14. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging*. 2016;35:1240–1251.
15. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Med Image Comput Assist Interv*. 2013;16(Pt 2):246–253.
16. Tran PV. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *ArXiv*. 2016;160400494.
17. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 4th International Conference on 3D Vision (3DV), Stanford, CA; 2016. pp 565–571.
18. Emad O, Yassine IA, Fahmy AS. Automatic localization of the left ventricle in cardiac MRI images using deep learning. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milano, Italy; 2015. pp 683–686.
19. Payer C, Štern D, Bischof H, Urschler M. Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham, Switzerland: Springer International Publishing; 2016:230–238.
20. Ulrich NH, Ahmadli U, Woernle CM, Alzarhani YA, Bertalanffy H, Kollias SS. Diffusion tensor imaging for anatomical localization

of cranial nerves and cranial nerve nuclei in pontine lesions: initial experiences with 3T-MRI. *J Clin Neurosci*. 2014;21:1924–1927.
21. Chen H, Shen C, Qin J, et al. Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham, Switzerland: Springer International Publishing; 2015. pp 515–522.
22. Rasti R, Teshnehlab M, Phung SL. Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. *Pattern Recognit*. 2017;72:381–390.
23. Hosseini-Asl E, Gimel'farb G, El-Baz A. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. *ArXiv*. 2016;160700556.
24. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18:1527–1554.
25. Liang X, Shen Z, Xiang D, Feng J, Lin L, Yan S. Semantic object parsing with local-global long short-term memory. *ArXiv*. 2015;151104510.
26. Zhang X, Lu L, Lapata M. Tree recurrent neural networks with application to language modeling. *ArXiv*. 2015.
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv*. 2014;14091556.
28. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–1780.
29. Zhao R, Ouyang W, Li H, Wang X. Saliency detection by multi-context deep learning. In Proceedings of the 28th Annual IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA; 2015: 1265–1274.
30. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik B, Swami A. The limitations of deep learning in adversarial settings. In Proceedings of the IEEE European Symposium on Security and Privacy (EuroS P), Saarbrücken, Germany; 2016: 372–387.
31. Kotikalapudi R. keras-vis. 2017. https://github.com/raghakot/keras-vis. Accessed January 23, 2019.
32. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *ArXiv*. 2015;151207951.
33. Zhang D, Icke I, Dogdas B, et al. A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, Washington DC; 2018: 470–473.
34. Xue W, Nachum IB, Pandey S, Warrington J, Leung S, Li S. Direct estimation of regional wall thicknesses via residual recurrent neural network. *ArXiv*. 2017;170509728.

How to cite this article: Bahrami N, Retson T, Blansit K, Wang K, Hsiao A. Automated selection of myocardial inversion time with a convolutional neural network: Spatial temporal ensemble myocardium inversion network (STEMI-NET). *Magn Reson Med*. 2019;00:1–9. https://doi.org/10.1002/mrm.27680