Impact of Diffusion–Perfusion Mismatch on Predicting Final Infarction Lesion Using Deep Learning

SUNGDONG LEE1, LEONARD SUNWOO2,3, YOUNGWON CHOI4,5, JAE HYUP JUNG2, SEUNG CHAI JUNG6, AND JOONG-HO WON1, (Member, IEEE)

1Department of Statistics, Seoul National University, Seoul 08826, Republic of Korea
2Department of Radiology, Seoul National University Bundang Hospital, Seongnam-si 13620, Republic of Korea
3Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam-si 13620, Republic of Korea
4Department of Radiological Sciences, University of California at Los Angeles, Los Angeles, CA 90095, USA
5Center for Computer Vision and Imaging Biomarkers, University of California at Los Angeles, Los Angeles, CA 90095, USA
6Department of Radiology, Asan Medical Center, Seoul 05505, Republic of Korea

Corresponding authors: Leonard Sunwoo (leonard.sunwoo@gmail.com) and Joong-Ho Won (woj@stats.snu.ac.kr)

This work was supported by the National Research Foundation of Korea under Grant NRF-2018R1C1B6007917 and Grant 2019R1A2C1007126.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Boards of Seoul National University Bundang Hospital under Application No. B-1707-408-103 and the Asan Medical Center under Application No. S2017-0807-0001.

ABSTRACT
We report a study that validates the impact of diffusion-perfusion mismatch in a deep learning (DL) model predicting the final infarction lesion from baseline magnetic resonance imaging (MRI). From 472 consecutive patients with acute ischemic stroke, we gathered baseline and follow-up MRI having intervals of 3–7 days, and initial and final infarction lesions were segmented. Four U-Net-based DL models from baseline MRI with different combinations of diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI) maps, and initial diffusion-restricted lesion prediction map (Pred\textsubscript{init} map) were trained to predict the final infarction lesion. Five-fold cross-validation was used for training and testing. As an external test set, 55 patients from another institution were analyzed. Dice similarity coefficient (DSC) was compared between the models and subgroups according to the presence of lesion growth and/or diffusion-perfusion mismatch. The model using the PWI maps and Pred\textsubscript{init} map showed the best mean DSC (0.422 and 0.486 for internal and external test set, respectively). This model showed better performance in predicting rapid lesion growth compared with the baseline model (mean DSC difference, 0.040; 95% confidence interval: 0.018–0.062). Using the PWI map with initial diffusion-restricted lesion prediction improved the performance of DL model in predicting the final infarction lesion from baseline MRI.

INDEX TERMS
Magnetic resonance imaging, diagnostic imaging, diffusion-perfusion mismatch, final infarction lesion, image segmentation, U-Net.

I. INTRODUCTION
Accurate and timely diagnosis is crucial for managing acute ischemic stroke. Particularly, assessing the extent of infarcted brain tissue and the time elapsed from stroke onset at baseline neuroimaging study is important in deciding the most appropriate treatment option. In addition, prediction of the final infarction lesion at baseline imaging is important since it is correlated with the functional outcome [1], [2].

Computed tomography (CT) and magnetic resonance imaging (MRI) are the mainstay of imaging modalities in diagnosing acute ischemic stroke. Although both imaging protocols can be used to assess large vessel occlusion and tissue perfusion status, the initial imaging strategy as to which modality (CT versus MRI) should be used first may differ across different institutions depending on their situations [3], [4].
For MRI, diffusion-weighted MR imaging (DWI) and dynamic susceptibility contrast perfusion-weighted MR imaging (PWI) are widely adopted tools for evaluating ischemic and oligemic regions in patients with acute ischemic stroke. A diffusion-restricted lesion on DWI indicates an infarcted tissue, and a tissue showing perfusion delay on PWI without diffusion restriction on DWI (i.e., DWI-PWI mismatch) is considered to be at risk but not yet infarcted, or ischemic penumbra [7], [8]. Similarly, the mismatch of visibility between DWI and fluid-attenuated inversion recovery (FLAIR) imaging (i.e., DWI-FLAIR mismatch) is known to be useful in estimating the lesion age, since it reflects the net increase of water during the transition of acute ischemic stroke to the subacute stage [9]. Thus, FLAIR, DWI, and PWI are essential in determining the severity of ischemic injury and extent of surrounding tissue at risk.

The recent advances in deep learning (DL) have shown promising results on various imaging modalities. Several studies have shown its effectiveness in the segmentation of acute ischemic stroke lesions on DWI [10], [11].

In this study, we aimed to develop a DL model to predict the final infarction lesion from the baseline (or initial) MRI in patients with acute ischemic stroke. Particularly, we aimed to evaluate the impact of DWI-PWI mismatch in enhancing the performance of final infarction lesion prediction as compared to that using FLAIR and DWI only. To assess our model’s generalizability, we performed external validation using data from another institution. In sum, the contributions of our work are as follows:

- We developed a DL model to predict the final infarction lesion from the initial MR imaging using the real-world dataset, composed of 472 consecutive patients with acute stroke, and performed external validation using additional data of 55 patients from another institution.
- The proposed model was able to predict the final infarction lesion with Dice similarity coefficient of 0.422–0.486.
- The addition of perfusion maps and the probability map of diffusion-restricted lesions on initial MR imaging significantly improved the model performance in predicting the final infarction lesion.

II. MATERIALS AND METHODS

This retrospective study was approved by the institutional review boards of two participating institutions, and informed consent was waived.

A. STUDY POPULATION

At Seoul National University Bundang Hospital (SNUBH), 1140 consecutive patients who underwent an initial MRI study from April 2014 to June 2017 were initially selected from the clinical database using the following inclusion criteria: 1) established diagnosis of acute ischemic stroke in patients aged ≥ 19 years; 2) time from symptom onset to initial MRI of < 24 h; 3) DWI, FLAIR, and PWI included in the initial MRI; and 4) follow-up MRI including DWI and FLAIR 3–7 days after the initial MRI. We excluded 668 patients due to the following conditions: 1) performance of mechanical thrombectomy prior to the follow-up MRI, 2) missing raw data of DWI or PWI required for image processing, and 3) failed automatic co-registration of initial and follow-up studies or inadequate image quality for interpretation. We excluded patients who underwent mechanical thrombectomy because the procedure directly affects the perfusion status, which would subsequently affect the prediction of final infarction lesion by the model (leading to underestimate the lesion growth when such data are used for training). Finally, 472 patients were enrolled in this study as training and internal test datasets. Out of the 668 excluded cases, for 208 patients whose DWI and FLAIR of the initial study were adequate, DWI and FLAIR of the initial study were used for training the DWI lesion segmentation model (Predinit). As an external test dataset, 55 patients who underwent initial MRI studies from January 2016 to December 2016 at Asan Medical Center (AMC) were enrolled using the same inclusion and exclusion criteria (Fig. 5). Although SNUBH followed a CT-based triaging system, AMC adopted an MRI-based triaging system.

The demographic and clinical data of the patients were collected including age, sex, risk factor of stroke such as hypertension, hyperlipidemia, or heart disease, stroke etiology, National Institutes of Health Stroke Scale rating, modified Rankin Scale (mRS), intravenous thrombolysis (IVT), and the time elapsed from the symptom onset to hospital visit (Table 5).

B. MRI EXAMINATION

At SNUBH, MRI scans were performed using 1.5T (Intera; Philips Healthcare, Best, the Netherlands) or 3T (Achieva or Ingenia; Philips Healthcare) MRI scanner. At AMC, all MRI scans were performed on a 1.5T scanner (Magnetom Avanto; Siemens Healthcare, Erlangen, Germany). Imaging sequences of the initial MRI included spin-echo echoplanar DWI, axial turbo spin-echo FLAIR, and gradient-echo echo-planar dynamic susceptibility contrast PWI, and the follow-up MRI included DWI and axial FLAIR. The apparent diffusion coefficient (ADC) map at the standard b-value (b = 1000 s/mm²) was generated from DWI on a voxel-by-voxel basis. Detailed MRI parameters were as follows: for DWI, repetition time (TR), 3000–8000 ms; echo time (TE), 56–103 ms; flip angle (FA), 90°; matrix, 128 × 128; field of view (FOV), 216 × 220–250 × 250 mm²; number of excitations, 1–6; slice thickness, 5 mm; and interslice gap, 1–2 mm; for FLAIR, TR, 9000–11200 ms; TE, 109–140 ms; FA, 90°; matrix, 256 × 162–352 × 264; FOV, 182 × 230–230 × 240 mm²; number of excitations, 1; slice thickness, 5 mm; and interslice gap, 1–2 mm; and for PWI, TR, 1470–1500 ms; TE, 29–40 ms; FA, 90°; matrix, 112 × 88–128 × 128; FOV, 216 × 230–230 × 240 mm²;
number of excitations, 1; slice thickness, 5 mm; and inter-
slice gap, 1–2 mm. For contrast enhancement, gadobutrol
(Gadovist®, Bayer Schering Pharma AG, Berlin, Germany;
0.1 mmol/kg) and gadoterate meglumine (Dotarem®; Guer-
bet, Aulnay-sous-Bois, France; 0.1 mmol/kg) were injected
as a bolus intravenously.

C. IMAGE ANALYSIS AND PROCESSING
Two board-certified neuroradiologists (L.S. and S.C.J.) with
11 and 16 years of clinical experience, respectively, reviewed
the MRIs and determined the reference standard or ground
truth for the final infarction lesion based on consensus.
FLAIR hyperintensity lesion on follow-up MRI corre-
sponding to acute ischemic lesion was considered as the final
infarction lesion. Areas of hemorrhagic transformation of
infarction were included in the infarction lesion. Defi-
nite subacute infarction on the initial MRI was excluded.
The ground truth masks were carefully drawn using
in-house software derived from ImageJ (U.S. National
Institutes of Health, Bethesda, Maryland, USA; available at
https://imagej.nih.gov/ij/).

Time-to-peak (TTP) and Tmax maps were generated
from PWI using the Olea software (Olea Medical Solu-
tions, La Ciotat, France). The Medical Imaging Pro-
cessing, Analysis, and Visualization (MIPAV) software
version 8.0.2 (Center of Information Technology and
National Institute of Health, Bethesda, Maryland, USA) was
used for the co-registration of the initial and follow-up MRI
examinations.

All pairs of MRI examinations were classified accord-
ing to presence of DWI-PWI mismatch and lesion growth.
DWI-PWI mismatch was defined as the mismatch ratio (area
of Tmax or TTP prolongation and that of diffusion-restricted
lesion on DWI) greater than 1.2 and absolute difference of
PWI and DWI abnormality greater than 10 mL [12]. Sim-
ilarly, lesion growth was defined as the ratio of the final
infarction lesion and initial diffusion-restricted lesion on DWI
greater than 1.2 and the final lesion size greater than 10 mL.

Before the training, we resized all images to the common
dimension of 20 × 512 × 512 (depth×height×width) voxels.
We also scaled the intensity level of the images to make all
values lie between 0 and 1, where extremely high values were
set to 1 and background noise was set to 0. Data augmentation
was not applied.

D. MODEL ARCHITECTURE AND TRAINING
To construct a possible prediction model on 3D images,
we developed a DL model that is a streamlined version of
the 3D multiscale residual U-Net [13], which won the first prize
in the Ischemic Stroke Lesion Segmentation (ISLES) Chal-
 lenges consecutively in 2016 and 2017 [14]. Our DL model,
whose architecture is described in the appendix (Table 6),
is simpler and uses less parameters than the original archi-
tecture while performing on par. Architectures based on
the U-Net [15], [16] have reported decent performance in
predicting lesion outcome with multi-spectral MRI despite
the difficulty of the task [14].

For the input images, we considered employing lesion
prediction map at the baseline (Pred\textsubscript{init} map) in addition
to diffusion- and perfusion-weighted maps. We separately
trained a segmentation model with the same architecture
(Table 6) using DWI, ADC, FLAIR, and the ground truth
maps at the same time point (i.e., the baseline lesion mask
was used for the baseline image), and the trained probability
maps obtained from this segmentation model were used as
an input of the prediction model. The inputs for the resulting
models (Models 1–4) are summarized in Table 1.

| Model No. | Input image sequences |
|-----------|-----------------------|
| 1         | DWI, ADC, FLAIR       |
| 2         | DWI, ADC, FLAIR, Tmax, TTP |
| 3         | DWI, ADC, FLAIR, Pred\textsubscript{init} |
| 4         | DWI, ADC, FLAIR, Tmax, TTP, Pred\textsubscript{init} |

For the loss function, we tried to directly optimize the
common criterion for segmentation quality known as Dice
similarity coefficient (DSC), defined as

\[
\text{DSC} = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}},
\]

where \(N_{TP}, N_{FP},\) and \(N_{FN}\) denote the number of true positive,
false positive, and false negative voxels predicted, respec-
respectively. By defining the loss function as the continuous version of
DSC,

\[
\ell(y, p) = \frac{2\sum_{i=0}^{N-1} y_i p_i}{[2\sum_{i=0}^{N-1} y_i] + \sum_{i=0}^{N-1} (1-y_i) p_i + \sum_{i=0}^{N-1} (1-p_i)},
\]

where \(y_i\) is the ground truth value (either 0 or 1) and \(p_i\)
is the predicted probability of \(i\)-th voxel, we can compute and
optimize the parameters of the DL model with respect to this
loss function. As we are trying to predict the lesion occurring
for only small regions on an image, DSC is more suitable
than other common losses such as the binary cross-entropy
or mean absolute error loss, because the former tends to favor
true positive predictions over true negatives.

We applied patch sampling for constructing the training
batch. The training batch comprised of 3D patches with a
size of 20 × 64 × 64 voxels, sampled from preprocessed
images. To accelerate the training process (or to save the
memory capacity demanded for the sampling), we randomly
selected at most 100 patients for every 100 iterations, and all
the patches for the training batch were sampled within those
selected patients for that 100 iterations. We optimized the
training loss using ADAM [17] for at most 20,000 iterations.
The early stopping method was applied by computing the
mean DSC of the validation set for every 100 iterations. Then,
the model parameters with the highest validation DSC was
selected.
We applied five-fold cross-validation (CV) by dividing all patients into five sets randomly, where three sets were used in training, one set was used for validation, and one set was used for testing for each fold. All models were trained with an identical split of the folds in a five-fold CV.

For the evaluation of the models, prediction maps about the final infarction lesion on test data were drawn out using a probability threshold of 0.50 for each voxel. Subsequently, DSC, sensitivity, and the positive predictive value (PPV) of the prediction maps were calculated. For SNUBH data, a test prediction map for each patient was obtained from the fold that was not used in training. For AMC data, prediction maps were obtained from majority voting among all five models obtained by five-fold CV.

The implementation of the model training and evaluation was based on Keras (version 2.3.1: https://keras.io) with TensorFlow (version 1.15.0; https://www.tensorflow.org) back-end. Using a workstation with Xeon® Processor E5-2650 version 4 (Intel, Santa Clara, CA, USA) and Titan X graphical processing units (Nvidia, Santa Clara, CA, USA), the computing time was shortest for Model 1 (40 h) and longest for Model 4 (65 h). The implemented training code can be found in https://github.com/sdlee087/deep_medical.

E. STATISTICAL ANALYSIS

Welch’s t-test was conducted to test the difference in the means of the numerical variables, and the chi-squared test was performed for categorical variables to test their homogeneity. We also applied correlation tests to analyze the effect of patients’ clinical information on the performance of these DL models. To compare the performance of the models, the DSC sensitivity and PPV of test prediction maps were compared using paired t-tests between the models. All analyses were processed with the statistical software R (version 3.6.3, R Foundation for Statistical Computing, Vienna, Austria).

III. RESULTS

The two participating institutions showed significant differences in several clinical features such as age, proportion of smokers, atrial fibrillation, coronary heart disease, sleep apnea, and proportion of 3T MRI scanner (Table 5). Approximately 30% and 45% of the patients from SNUBH and AMC, respectively, exhibited DWI-PWI mismatch. In addition, the initial and final infarct volumes of patients from AMC (median: 17.8 mL; interquartile range [IQR]: 3.0–44.8 mL; and median: 22.1 mL; IQR: 4.3–71.8 mL, respectively) were significantly higher than those from SNUBH (median: 0.86 mL; IQR: 0.33–3.44 mL; and median: 1.47 mL; IQR: 0.52–6.67 mL, respectively).

The loss curve of the trained model is presented in Fig. 6. We could notice that the models with Pred\textsubscript{init} (Model 3, Model 4) achieves optimal weight quicker than other models. The test measures of each model are presented in Table 2. Among all models, Model 4 showed the highest DSC for both internal and external test sets (0.422 ± 0.277 [standard deviation] and 0.486 ± 0.252, respectively). See Section IV for further discussions on the impact of Pred\textsubscript{init} to the model.

Compared with the baseline model (Model 1), Model 4 showed a significant difference in mean DSC in total data under 5% significance level. In terms of subgroup analysis, there was a statistically significant difference between two models in subgroups such as internal and external test sets, data with presence of lesion growth, data with DWI-PWI mismatch, and DWI-PWI mismatch with lesion growth. Particularly, the mean DSC increment of the subgroup with lesion growth was 0.040 (95% confidence interval: 0.018–0.062) (Fig. 1). In addition, we could report that the volume of the lesion significantly affects the accuracy of the prediction map; for bigger lesions, the prediction maps from models were more likely to achieve higher DSC, which increases the mean and decreases the variance of the test DSC (Fig. 2). For both models, mean test DSCs were significantly higher in subjects with lesion volume ≥ 10 mL than in those with lesion volume < 10 mL (p < 0.001 for both) on the result of the internal test set. However, the external test set did not show such a significant difference for both models due to the small sample size and high variance in the test DSC (Table 3).

![FIGURE 1. Mean and 95% confidence interval of DSC difference (Model 4 - Model 1) on various subgroups. The intervals above the red dashed line indicate results with statistical significance. DSC = dice similarity coefficient. LG = lesion growth. DPM = DWI-PWI mismatch.](image-url)

The results of the correlation tests between mean DSC and DWI-PWI mismatch or the rate of lesion growth in Model 1, Model 2, and Model 4 are shown in Table 4. For Model 4, DWI-PWI mismatch and lesion growth (growth more than 20%) features were negatively correlated with the test DSC (\(\hat{\rho} = -0.134\) and \(-0.225\), respectively; \(p = 0.002\) and 0.011, respectively). For Model 1, DWI-PWI mismatch exhibited a significant negative correlation with the DSC (\(\hat{\rho} = -0.086, p = 0.049\)), but the effect of lesion growth was uncertain (\(p > 0.05\)). For Model 2, neither
TABLE 3. Mean (standard deviation) of test DSC of Model 1 and Model 4 by lesion volume. The mean (standard deviation) of test DSC for each subgroup is given. The result of Welch’s t-test between the statistics of each subgroup is given as in p-value.

| Data          | Statistic | Subgroup by final lesion volume | p-value |
|---------------|-----------|---------------------------------|---------|
|               |           | < 10mL | > 10mL |         |          |
| Internal test set | DSC (Pred_{\text{init}}) | 0.460 (0.304) | 0.649 (0.263) | < 0.001* |         |
|               | DSC (Model 1) | 0.355 (0.281) | 0.581 (0.244) | < 0.001* |         |
|               | DSC (Model 4) | 0.377 (0.269) | 0.607 (0.230) | < 0.001* |         |
| External test set | DSC (Pred_{\text{init}}) | 0.409 (0.264) | 0.486 (0.262) | 0.299    |         |
|               | DSC (Model 1) | 0.409 (0.276) | 0.419 (0.283) | 0.845    |         |
|               | DSC (Model 4) | 0.410 (0.253) | 0.533 (0.243) | 0.083    |         |

There were 26 cases in which the acute infarction lesion was not detected from the initial MRI scan (i.e., the initial lesion volume is zero), but subsequently developed on the follow-up MRI (internal test set, 25 cases; external test set, 1 case). Among these, Model 4 was able to predict the occurrence of acute infarction lesions in 5 cases (19.2%; internal test set, 4 cases; external test set, 1 case). When all models were considered, at least one model was able to predict the lesion occurrence in 8 cases (30.8%; internal test set, 7 cases; external test set, 1 case) (Fig. 4).

There were 26 cases in which the acute infarction lesion was not detected from the initial MRI scan (i.e., the initial lesion volume is zero), but subsequently developed on the follow-up MRI (internal test set, 25 cases; external test set, 1 case). Among these, Model 4 was able to predict the occurrence of acute infarction lesions in 5 cases (19.2%; internal test set, 4 cases; external test set, 1 case). When all models were considered, at least one model was able to predict the lesion occurrence in 8 cases (30.8%; internal test set, 7 cases; external test set, 1 case) (Fig. 4).

There were 26 cases in which the acute infarction lesion was not detected from the initial MRI scan (i.e., the initial lesion volume is zero), but subsequently developed on the follow-up MRI (internal test set, 25 cases; external test set, 1 case). Among these, Model 4 was able to predict the occurrence of acute infarction lesions in 5 cases (19.2%; internal test set, 4 cases; external test set, 1 case). When all models were considered, at least one model was able to predict the lesion occurrence in 8 cases (30.8%; internal test set, 7 cases; external test set, 1 case) (Fig. 4).

IV. DISCUSSION

The main focus of our study was to generate an algorithm to predict the final infarction lesion from the initial MR imaging using the real-world dataset. To that end, we trained a DL model with a U-Net architecture using the consecutive dataset consisting of 472 patients with acute stroke from a tertiary stroke center. In addition, we performed external validation using the data from another large institution. The results of our study show that the addition of perfusion maps and the...
probability map of diffusion-restricted lesions on initial MRI (Pred_initial map) improve the performance of predicting the final infarction lesion. Although the model was trained using the data from an institution with a CT-based triage system, this trend was confirmed using the dataset from another institution with a MR-based triage system. This suggests that the proposed model can be used in both institutions with CT-based and MR-based triaging systems.

Many studies have applied machine learning or DL approaches for stroke lesion segmentation [10], [11], [18], [19]. Zhang et al. [19] have achieved a DSC of 0.79 using a dataset of DWI and ADC from 242 patients. In a study by Kim et al. [11], overall mean DSC was 0.6. More specifically, the DSC was > 0.75 for lesion volume > 70 mL, whereas the DSC showed high variance for lesion volume < 25 mL. In this study, the mean DSC of segmentation task for the initial lesion on the internal test set was 0.497 (Table 2). Considering that our data has relatively small lesion volume, this result is in line with that of Kim et al. [11], indicating that smaller lesions tend to show a wide distribution of the DSC. On the other hand, in the subgroup having final lesion volume larger than 10 mL, the mean DSC of initial lesion segmentation achieved up to 0.649 (Table 3), which is comparable with previous studies.

In contrast, few studies have applied DL in predicting the final infarction lesion from the initial MRI. The task of the International Stroke Lesion Segmentation (ISLES) 2017 challenge [14] requested predicting the 90-day stroke lesion with only 43 cases as the training dataset, which is a significantly harder task compared to our study. The best model, similar to ours, was based on a U-Net architecture with a mean test DSC of 0.31. Recently, Yu et al. [20] have predicted final ischemic stroke lesions with attention-gated U-Net and fivefold CV and obtained a median DSC of 0.53. Although our result has a lower DSC, we must acknowledge that the lesion size of the dataset used on both studies is relatively different. Yu et al. [20] have reported the median lesion volume of the dataset of 182 patients as 54 mL (IQR: 16–117 mL), whereas the median volume of the final lesion in the internal test set is 1.47 mL (IQR: 0.52–6.67 mL). Considering the mean DSC of 0.649 and for lesion volume larger than 10 mL in the internal dataset used in our study (Table 3), we believe that the performance is comparable between the two studies.

In this study, we consecutively enrolled a relatively large number of acute ischemic stroke patients for training the DL models and validating their performance. A noteworthy aspect of our study is that we used an external test set by bringing in the data from another institution into these models. Whereas the treatment decisions at SNUBH are based on CT, those at AMC are based on MRI. As such, the proportion of large territorial infarctions differs in the two datasets, and small infarctions tended to be more frequent in the SNUBH dataset, which might have affected the prediction pattern in the AMC dataset. However, for both types of test sets, the performance of models using PWI maps was significantly higher than that only using DWI and FLAIR maps. In addition, the setting of this study may reflect the situation of the real world that different triaging methods are currently used across the hospitals.

It is clinically known that DWI-PWI and DWI-FLAIR mismatches are useful in determining the time point of infarction and possibility of the lesion growth [3], [6], but such predictions from analysis of the mismatches are still subjective, and the presence of DWI-PWI mismatch does not always identify lesion growth [21]. In this regard, our proposed method can aid in deciding appropriate treatment options by enabling accurate and timely lesion prediction. To highlight the effects of DWI-PWI mismatch, we adopted a probability map for initial lesion from a segmentation model (Pred_initial map). As a result, although a mere addition of PWI (Model 2) failed to show improvement of DSC, Model 4 (which is composed of DWI, ADC, FLAIR, Tmax, TTP, and Pred_initial map) showed a significantly higher DSC compared to other models. Thus, it seems that the combination of Pred_initial map and PWI maps led the network to pay more attention to DWI-PWI mismatch.

In a recent study by Lin et al. [22], the velocity of infarction lesion growth was found to be associated with the therapeutic benefit of mechanical thrombectomy. More specifically, they found that mechanical thrombectomy increased the odds of good clinical outcome for patients with rapid lesion growth of 25 mL/h. Since our proposed models are designed to predict the final infarction lesion based on initial MRI, we believe that our models, after some modulation, might aid in assessing the clinical outcome by predicting the lesion growth velocity.

Interestingly, we found that the proposed models were able to predict the occurrence of the final infarction lesion for approximately 30% of the cases where the lesion only appeared at the follow-up study and not at the initial study (Fig. 4). Hence, a future study to confirm the utility of the proposed method in distinguishing transient ischemic attack and actual infarction might be helpful.

Of note, the mean DSC of Model 2 (composed of DWI, ADC, FLAIR, TTP, and Tmax maps) of the external test set was lower than the internal test set, whereas the mean DSCs of all the other models of the external test set were higher than the internal test set (Table 5). We hypothesize that this discrepancy may stem from the different magnetic strength of the MRI scanners in the two institutions (Table 5), provoking mispredictions on external validation particularly for the PWI data. Nevertheless, Model 4 demonstrated consistent results, suggesting that using the Pred_initial map may help alleviate this domain effect between PWI maps originating from different MRI scanners.

This study has several limitations. First, the datasets used in this study were retrospectively collected and may not be sufficiently large to address the variability in scanning protocols and hardware implementations across the institutions. Nevertheless, we enrolled consecutive patients with acute ischemic stroke, which outnumber the training dataset of the ISLES 2017 by more than 10-fold, even though the
endpoints are different (3–7 days vs. 90 days). Additionally, we verified our model in the external test set and found consistent results. Second, follow-up MRIs were performed 3–7 days after initial imaging, based on the routine protocol of the two institutions, and served as the ground truth of the final infarction lesion in this study. Although such a time interval may not be optimal to judge the final infarction lesion, many related studies used similar intervals [23], [24], [25]. Finally, we excluded patients who underwent mechanical thrombectomy because the procedure significantly affects the perfusion status and lesion growth. However, since 30-45% of the patients from two participating institutions exhibited DWI-PWI mismatch, we believe that the contribution of DWI-PWI mismatch to lesion growth would have been at least partially learned by the model. Nevertheless, we acknowledge that the exclusion of patients who underwent mechanical thrombectomy may lead to selection bias regarding stroke type and severity. A future study including a larger number of patients undergoing mechanical thrombectomy may provide more valuable information in predicting the final infarct lesion as to whether the procedure would be beneficial.

V. CONCLUSION
In conclusion, the proposed DL model showed potential in predicting the final infarction lesion from the initial MRI. The model with combination of perfusion MRI and initial diffusion-restricted lesion in addition to DWI and FLAIR showed the best performance.

APPENDIX A
ADDITIONAL TABLES AND FIGURES
Figure 5 summarizes the procedure on collecting the study population. Table 5 summarizes overall statistics of the study population. Table 6 summarizes the train architecture.
TABLE 6. Detailed structure of the deep learning architecture. Batch normalization were applied after the convolutional operation of each layer except the last layer. All convolutional layers used in the architecture were padded with “same” mode and initialized with “he_normal” mode [26] supported in Keras. All models in this study follow this structure, and only input shape differs by the number of input images.

| Layer | Operation | Filters | Kernel | Strides | Activation | Linked layer | Output Shape |
|-------|-----------|---------|--------|---------|------------|--------------|--------------|
| Input |           |         |        |         |            |              | [20, 512, 512, k] |
| 1     | Conv      | 16      | 1x1    | 1x1     | ReLU       | Input        | [20, 256, 256, 16] |
| 2     | Conv      | 16      | 1x3x3  | 1x1     | ReLU       | 1            | [20, 256, 256, 16] |
| 3-1   | Conv      | 32      | 1x1x1  | 1x1x1   | ReLU       | 2            | [20, 128, 128, 32] |
| 3-2   | Conv      | 2       | 1x1    | 1x1     | ReLU       | 2            | [20, 256, 256, 2] |
| 4     | Conv      | 16      | 1x1x1  | 1x1     | ReLU       | 3-1          | [20, 128, 128, 16] |
| 5-1   | Conv      | 32      | 1x1x3  | 1x2x2   | ReLU       | 4            | [20, 64, 64, 32] |
| 5-2   | Conv      | 4       | 1x1x1  | 1x1     | ReLU       | 4            | [20, 128, 128, 4] |
| 6     | Conv      | 32      | 3x3x3  | 1x1x1   | ReLU       | 5-1          | [20, 64, 64, 32] |
| 7     | T. Conv   | 16      | 1x1x3  | 1x2x2   | ReLU       | 6            | [20, 128, 128, 16] |
| 8     | T. Conv   | 16      | 1x3x3  | 1x1x1   | ReLU       | 5-2, 7       | [20, 128, 128, 16] |
| 9     | T. Conv   | 16      | 1x3x3  | 1x2x2   | ReLU       | 8            | [20, 256, 256, 16] |
| 10    | T. Conv   | 8       | 1x1x3  | 1x2x2   | ReLU       | 3-2, 9       | [20, 512, 512, 8] |
| Output| Conv      | 1       | 3x5x5  | 1x1x1   | Sigmoid    | Input, 10    | [20, 512, 512, 1] |

ACKNOWLEDGMENT

The authors acknowledge the assistance of Minkyung Kang and Green Lee.

REFERENCES

[1] A. J. Yoo, Z. A. Chaudhry, R. G. Nogueira, M. H. Levy, P. W. Schaefter, L. H. Schwamm, J. A. Hirsch, and R. G. González, “Infarct volume is a pivotal biomarker after intra-arterial stroke therapy,” Stroke, vol. 43, no. 5, pp. 1323–1330, May 2012.

[2] S. F. Zaidi, A. Aghaebrahim, X. Urra, M. A. Jumaa, B. Jankowitz, M. Hammer, R. Nogueira, M. Horowitz, V. Reddy, and T. G. Jovin, “Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy.” Stroke, vol. 43, no. 12, pp. 3238–3244, Dec. 2012.

[3] T. Shang and D. R. Yavagal, “Application of acute stroke imaging: Selecting patients for revascularization therapy,” Neurology, vol. 79, no. 13, pp. S86–S94, Sep. 2012.

[4] O. Y. Bang, J.-W. Chung, J. P. Son, W.-S. Ryu, D.-E. Kim, W.-K. Seo, G.-M. Kim, and Y.-C. Kim, “Multimodal MRI-based triage for acute stroke therapy: Challenges and progress,” Frontiers Neurol., vol. 9, p. 586, Jul. 2018.

[5] G. Schlau, A. Benfield, A. Baird, B. Siewert, K. Lövblad, R. Parker, R. Edelman, and S. Warach, “The ischemic penumbra: Operationally defined by diffusion and perfusion MRI,” Neuroradiology, vol. 53, no. 7, p. 1528, 1999.

[6] W.-D. Heiss, “The ischemic penumbra: Correlates in imaging and implications for treatment of ischemic stroke,” Cerebrovascular Diseases, vol. 32, no. 4, pp. 307–320, 2011.

[7] G. W. Albers, M. P. Marks, S. Kemp, S. Christensen, J. P. Tsai, S. Ortega-Gutierrez, R. A. McTaggart, M. T. Torrey, M. Kim-Tenser, and T. Leslie-Mazwi, “Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging.” New England J. Med., vol. 378, no. 8, pp. 708–718, 2018.

[8] M. Bouslama, D. C. Haussen, J. A. Grossberg, S. Dehkargarli, M. T. Bowen, L. C. Rebello, N. A. Bianchi, M. R. Frankel, and R. G. Nogueira, “Computed tomographic perfusion selection and clinical outcomes after endovascular therapy in large vessel occlusion stroke,” Stroke, vol. 48, no. 5, pp. 1271–1277, May 2017.

[9] M. Tourniaire, B. Cheng, M. Ebinger, O. Hao, T. Tourdias, W. J. Kim, L. Breuer, O. C. Singer, S. Warach, and S. Christensen, “DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4 - 5 h of symptom onset (PRE-FLAIR): A multicentre observational study,” Lancet Neurol., vol. 10, no. 11, pp. 978–986, 2011.

[10] I. Woo, A. Lee, S. C. Jung, H. Lee, N. Kim, S. J. Cho, D. Kim, J. Lee, L. Sunwoo, and D.-W. Kang, “Fully automatic segmentation of acute ischemic lesions on diffusion-weighted imaging using convolutional neural networks: Comparison with conventional algorithms,” Korean J. Radiol., vol. 20, no. 8, p. 1275, 2019.

[11] Y.-C. Kim, J.-E. Lee, I. Yu, H.-N. Song, I.-Y. Baek, J.-K. Seong, H.-G. Jeong, B. J. Kim, H. S. Nam, J.-W. Chung, O. Y. Bang, G.-M. Kim, and W.-K. Seo, “Evaluation of diffusion lesion volume measurements in acute ischemic stroke using encoder–decoder convolutional network,” Stroke, vol. 50, no. 6, pp. 1444–1451, Jun. 2019.

[12] S. M. Davis, “Effects of alteplase beyond 3 h after stroke in the echoplanar imaging thrombolytic evaluation trial (EPITHET): A placebo-controlled randomised trial,” Lancet Neurol., vol. 7, no. 4, pp. 299–309, Apr. 2008.

[13] Y. Choi, Y. Kwon, H. Lee, B. J. Kim, M. C. Paik, and J.-H. Won, “Ensemble of deep convolutional neural networks for prognosis of ischemic stroke,” in Proc. Int. Workshop Brainlesion, Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries. Cham, Switzerland: Springer, 2016, pp. 231–243.

[14] S. Winzeck, “Isles 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI,” Frontiers Neurol., vol. 9, p. 679, Sep. 2018.

[15] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI). Munich, Germany, 2015, pp. 234–241.
S. Lee et al.: Impact of Diffusion–Perfusion Mismatch on Predicting Final Infarction Lesion Using DL

[16] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2016, pp. 424–432.

[17] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

[18] L. Chen, P. Bentley, and D. Rueckert, “Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks,” NeuroImage, Clin., vol. 15, pp. 633–643, Jun. 2017.

[19] R. Zhang, L. Zhao, W. Lou, J. M. Abrego, V. C. Mok, W. C. Chu, D. Wang, and L. Shi, “Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densetests,” IEEE Trans. Med. Imag., vol. 37, no. 9, pp. 2149–2160, Sep. 2018.

[20] Y. Yu, Y. Xie, T. Thamm, E. Gong, J. Ouyang, C. Huang, S. Christensen, M. P. Marks, M. G. Lansberg, G. W. Albers, and G. Zaharchuk, “Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging,” JAMA Neurol. Open, vol. 3, no. 3, Mar. 2020, Art. no. e200772.

[21] C. S. Rivers, J. M. Wardlaw, P. A. Armitage, M. E. Bastin, T. K. Carpenter, V. Cvoor, P. J. Hand, and M. S. Dennis, “Do acute diffusion- and perfusion-weighted MRI lesions identify final infarct volume in ischemic stroke?” Stroke, vol. 37, no. 1, pp. 98–104, Jan. 2006.

[22] L. Lin, H. Zhang, C. Chen, A. Bivard, K. Butcher, C. Garcia-Esparon, N. J. Sratat, C. R. Levi, M. W. Parsons, and G. Li, “Stroke patients with faster core growth have greater benefit from endovascular therapy,” Stroke, vol. 52, no. 12, pp. 3998–4006, Dec. 2021.

[23] T. Tordius, P. Renou, I. Sibon, J. Asselineau, L. Bracoud, M. Dumoulin, F. Rouanet, J. Orgogozo, and V. Dousset, “Final cerebral infarct volume is predictable by MR imaging at 1 week,” AJNR. Amer. J. Neuroradiol., vol. 32, no. 2, pp. 352–358, 2011.

[24] H. M. Wheeler, M. Mlynash, M. Inoue, A. Tipirneni, J. Liggins, G. Zaharchuk, M. Straka, M. G. Lansberg, and G. W. Albers, “Early diffusion-weighted imaging and perfusion-weighted imaging lesion volumes forecast final infarct size in DEFUSE 2,” Stroke, vol. 44, no. 3, pp. 681–685, Mar. 2013.

[25] G. W. Albers, “Ischemic core and hyperperfusion volumes predict infarct size in SWIFT PRIME,” Ann. Neurol., vol. 79, no. 1, pp. 76–89, Jan. 2016.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1026–1034.

SUNDONG LEE received the B.S. degree in statistics and computational science from Seoul National University, South Korea, where he is currently pursuing the Ph.D. degree with the Department of Statistics. Since 2017, he has been a Research Assistant with the Computational Statistics Research Group, Department of Statistics, Seoul National University. His research interests include deep models for analyses of image data, deep generative models, and high-performance statistical AI. He was a recipient of the Korean Statistical Society Summer Conference Award of Excellence for Graduate Students, in 2022.

LEONARD SUNWOO received the M.D., M.S., and Ph.D. degrees in medicine from Seoul National University, Seoul, South Korea, in 2006, 2014, and 2018, respectively. From 2010 to 2014, he trained at the Department of Radiology, Seoul National University Hospital, Seoul, during his residency. He also completed a fellowship at Seoul National University Hospital, Seoul, during his residency. He is currently working as an Associate Professor at the Department of Radiology, Seoul National University Bundang Hospital, Seongnam-si, South Korea, since 2016. His research interests include the machine learning and deep learning applications in medical imaging, brain tumor, and stroke imaging. He is a member of the Korean Society of Radiology, Korean Society of Neuroradiology, Korean Society of Magnetic Resonance in Medicine, International Society for Magnetic Resonance in Medicine, and Radiological Society of North America.

JOONG-HO WON (Member, IEEE) received the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, in 2009. He is currently a Professor in statistics at Seoul National University. His research interests include high-performance statistical computing, convex and nonconvex optimization, machine learning, and image processing.

YOUNGWON CHOI received the B.S. degree in physics, the M.S. degree in the interdisciplinary program in computational science and technology, and the Ph.D. degree in statistics from Seoul National University, Seoul, South Korea, in 2012, 2015, and 2020, respectively. From 2016 to 2020, she was a Research Assistant with the Computational Statistics Research Group, Department of Statistics, Seoul National University. Since 2021, she has been a Postdoctoral Fellow with the Center for Computer Vision & Imaging Biomarkers, Department of Radiological Science, University of California at Los Angeles, Los Angeles. Her research interests include translating AI to clinical practice, including explainable AI, domain generalization, medical applications with human-AI interaction, and deep generative models for robust representation learning. Her awards and honors include the cum laude and honorable mention in computer-aided diagnosis poster awards from the SPIE Medical Imaging Conference, and the 2nd prize in open-source software demonstration from the SPIE Medical Imaging Conference. She received two first prizes in the MICCAI ISLES segmentation challenge.

JAE HYUP JUNG received the M.D. degree in medicine from Seoul National University, Seoul, South Korea, in 2018. From 2018 to 2019, he finished medical internship at Armed Forces Capital Hospital, Seongnam-si, South Korea. Currently, he is training the residency with the Department of Radiology, Seoul National University Bundang Hospital, Seongnam-si. He is a member of the Korean Society of Radiology.

SEUNG CHAI JUNG received the M.D., M.S., and Ph.D. degrees in medicine from Seoul National University, Seoul, South Korea, in 2004, 2009, and 2015, respectively. From 2004 to 2014, he trained at the Department of Radiology, Seoul National University Hospital, Seoul, during his residency. He also completed a fellowship at Seoul National University Hospital. Since 2014, he has been working as an Assistant Professor with the Asan Medical Center, University of Ulsan College of Medicine, Seoul, where he is currently working as an Associate Professor at the Department of Radiology, Asan Medical Center. His research interests include the stroke and vascular imaging in neuroradiology. He is a member of the Korean Society of Radiology, Korean Society of Neuroradiology, and Korean Society of Magnetic Resonance in Medicine.

VOLUME 10, 2022
97887

97887