NEUR**O**CRITICAL
CARE SOCIETY

ORIGINAL WORK

Machine Learning for Early Detection of Hypoxic-Ischemic Brain Injury After Cardiac Arrest

Ali Mansour[1,2†], Jordan D. Fuhrman[3†], Faten El Ammar[1], Andrea Loggini[1], Jared Davis[1], Christos Lazaridis[1,2], Christopher Kramer[1,2], Fernando D. Goldenberg[1,2*] and Maryellen L. Giger[3*]

Abstract

Background: Establishing whether a patient who survived a cardiac arrest has suffered hypoxic-ischemic brain injury (HIBI) shortly after return of spontaneous circulation (ROSC) can be of paramount importance for informing families and identifying patients who may benefit the most from neuroprotective therapies. We hypothesize that using deep transfer learning on normal-appearing findings on head computed tomography (HCT) scans performed after ROSC would allow us to identify early evidence of HIBI.

Methods: We analyzed 54 adult comatose survivors of cardiac arrest for whom both an initial HCT scan, done early after ROSC, and a follow-up HCT scan were available. The initial HCT scan of each included patient was read as normal by a board-certified neuroradiologist. Deep transfer learning was used to evaluate the initial HCT scan and predict progression of HIBI on the follow-up HCT scan. A naive set of 16 additional patients were used for external validation of the model.

Results: The median age (interquartile range) of our cohort was 61 (16) years, and 25 (46%) patients were female. Although findings of all initial HCT scans appeared normal, follow-up HCT scans showed signs of HIBI in 29 (54%) patients (computed tomography progression). Evaluating the first HCT scan with deep transfer learning accurately predicted progression to HIBI. The deep learning score was the most significant predictor of progression (area under the receiver operating characteristic curve = 0.96 [95% confidence interval 0.91–1.00]), with a deep learning score of 0.494 having a sensitivity of 1.00, specificity of 0.88, accuracy of 0.94, and positive predictive value of 0.91. An additional assessment of an independent test set confirmed high performance (area under the receiver operating characteristic curve = 0.90 [95% confidence interval 0.74–1.00]).

Conclusions: Deep transfer learning used to evaluate normal-appearing findings on HCT scans obtained early after ROSC in comatose survivors of cardiac arrest accurately identifies patients who progress to show radiographic evidence of HIBI on follow-up HCT scans.

*Correspondence: fgoldenb@neurology.bsd.uchicago.edu; m-giger@
uchicago.edu
†Ali Mansour and Jordan D. Fuhrman contributed equally to the first
authorship of this article.
[1] Neurosciences Intensive Care Unit, Department of Neurology, University
of Chicago Medicine and Biological Sciences, 5841 S. Maryland Ave., MC
2030, Chicago, IL 60637-1470, USA
[3] Department of Radiology, University of Chicago, 5841 S. Maryland Ave.,
Chicago, IL 60637-1470, USA
Full list of author information is available at the end of the article

Introduction

Despite the many advances in the field of resuscitation post cardiac arrest, hypoxic-ischemic brain injury (HIBI) leaves many survivors with severe neurological disability [1–4].

Patients who remain comatose at 24–72 h or more after resuscitation undergo neuroprognostication, aimed to detect signs of HIBI and predict projected long-term neurological function. Four main categories of tests are used to aid in achieving this purpose: clinical examination, electrophysiology, biomarkers, and neuroimaging [5]. The time at which these tests are done and their association with outcome also vary. Although clinical findings, such as pupillary reflexes and somatosensory evoked potentials, remain the most robust tests associated with outcome, biomarkers, electroencephalography, and imaging have inconsistencies that make interpreting them more dubious and subject to caution. Therefore, a multimodal approach to prognostication that factors in multiple tests is often recommended [6, 7].

The radiographic hallmark of HIBI is cerebral edema, commonly evaluated on head computed tomography (HCT) as effacement of sulci and diminished gray–white matter differentiation in cortical and deep brain structures. Although magnetic resonance imaging (MRI) is arguably a more accurate modality with higher resolution and capacity at identifying HIBI, it is not always feasible, and HCT remains the more prevalent and accessible neuroimaging modality for this patient population. Cerebral edema is often a progressive phenomenon that may or may not be appreciated by expert radiologists on initial HCT. Although there is presently no consensus on the optimal timing for performing brain computed tomography (CT) for neuroprognostication, most studies evaluate imaging done within the first 24 h, in which sensitivity of detecting cerebral edema is as low as 14% and increases to approximately 60% between 24 h and 7 days [7, 8]. Absolute decrease in, difference between, and ratio between gray and white matter density (GWM ratio) have been investigated in relation to neurological outcome. Although an absolute decrease in gray matter density alone is an unreliable predictor of poor neurological outcome, the predictive value increases when GWM ratio is considered. The main limitations to those studies include the time at which the analyzed HCT scan was performed (time from return of spontaneous circulation [ROSC] to HCT ranges anywhere from 4 to 72 h), and the choice of specific regions of interest within the brain to compare gray and white matter [9–12].

Recently, developments in machine learning considerably improved automatic execution of computer vision tasks in medical imaging, including disease detection, diagnosis, and segmentation [9]. Notably, convolutional neural networks—a family of deep learning architectures that identify desirable image features through optimized convolutional filters—perform comparably to experienced radiologists, with the added benefits of higher reading speeds and consistency. However, these models generally require a large amount of training data, often unavailable in medical imaging [13]. Therefore, many deep learning schemes have been developed in an attempt to circumvent this obstacle. One such technique is transfer learning, or the use of a model pretrained to perform a task in one domain that is then applied to a new domain [9, 13–16]. This approach preserves features that are useful for classification of the original image domain to classify images in the transferred domain while minimizing the need for new training data.

We hypothesize that progression of HIBI identified on follow-up HCT is in fact readily identifiable on an early initial scan and that the lower sensitivity observed within the first 24 h is more likely attributable to the subtle changes that evade the detection threshold of the human eye. As such, in a cohort of comatose survivors of cardiac arrest with reportedly normal HCT findings on presentation, we use machine learning to predict progression or not of HIBI only on the basis of the very first HCT scan.

Methods

Study COHORT

This is a case–control analysis of patients who suffered cardiac arrest, be it in the hospital or outside the hospital, in the time period between October 2017 and March 2020. The Institutional Review Board of the University of Chicago approved the protocol (IRB 200,107). For this type of study, formal consent was not required. Inclusion criteria for the study were as follows: (1) presenting diagnosis of cardiac arrest, (2) age ≥ 18 years, (3) unresponsive (comatose status) after ROSC, (4) noncontrast HCT imaging performed within 24 h of admission and deemed normal with no stigmata of HIBI by a board-certified neuroradiologist (in particular, no evidence of sulcal effacement, loss of gray–white matter differentiation, or compromise of cisternal spaces), and (5) available repeat HCT imaging within 2 to 10 days from the initial HCT scan. The following were the exclusion criteria: (1) dead on arrival, (2) failure to achieve ROSC, and (3) absence of HCT imaging within 24 h from arrest or absence of

follow-up HCT imaging within 10 days from that time. Although all patients had an initial HCT scan that was interpreted as lacking any signs of HIBI by a board-certified neuroradiologist, cases were defined as patients who suffered development of HIBI on repeat imaging and controls were patients who did not develop HIBI and continued to be interpreted as having no signs of HIBI. HIBI was defined on imaging as any evidence of sulcal effacement, loss of gray–white matter differentiation, or compromise of cisternal spaces.

Data Collection

For each patient, data regarding demographics, clinical presentation, Glasgow Coma Scale (GCS) scores, HCT scans on admission and follow-up, time intervals from presentation to initial imaging, time interval between HCT scans, laboratory studies, hospital length of stay, and discharge disposition were reviewed.

HCT images were reviewed by a board-certified neuroradiologist. Patients were categorized into two groups according to radiological reports of their follow-up HCT imaging. The first group included patients whose presenting HCT imaging results were evaluated and deemed lacking any signs of HIBI and whose follow-up imaging maintained that status (no CT progression [NCTProg]). The second group included patients whose presenting HCT imaging results were also deemed lacking any signs of HIBI; however, the follow-up imaging results were deemed as showing signs of HIBI (CT progression [CTProg]). This was a retrospective evaluation of reports by neuroradiologists. No specific instructions regarding imaging review windows were dictated. Furthermore, reports commenting on chronic findings, such as stigmata of small vessel disease, chronic subdural collections, atrophy, or prior surgical interventions, were not factored as abnormalities. The purpose of the aforementioned was to best depict real-life practice and to not bias readers by the purpose of the current study.

Deep Transfer Learning

CT scans were windowed with a standard brain window with center 40 HU and width 40 HU, and CT slices presenting no brain anatomy were excluded from analysis. The transfer learning approach used in this study was based on methods described by Antropova et al. [13], which were expanded to account for the three-dimensional information available in CT scans. Briefly, a VGG19 network architecture pretrained on the ImageNet database (a collection of millions of natural nonmedical images) was used to extract quantitative features from only the initial HCT scan (no follow-up information included), as described in Fig. 1 [17]. The mean value of each feature map produced by the maximum pooling layers was used to form a normalized representative feature vector for each individual CT slice. These vectors were maximum pooled in the axial dimension for all slices within a scan to obtain a scan-level representation.

Because of the limited data in this study, leave-one-out-by-patient cross-validation was used. Principal component analysis was performed on 53 of the 54 available scans for dimensionality reduction in an attempt to alleviate the risk of model overfitting. Then a support vector machine (SVM) was trained by using the principal components for the task of classifying a scan as progressive (exhibited or would exhibit signs of HIBI on follow-up HCT scan) or nonprogressive (no signs of HIBI on follow-up HCT scan). The single scan that was not included in the training set was then evaluated by using the SVM. This process of principal component analysis, SVM training, and single-scan testing was repeated so that each of the 54 scans served as the test scan exactly one time, and a prevalence scaling factor was applied to correct for class imbalance in the data set [18, 19]. This full workflow is depicted in Figs. 1 and 2, with the predicted SVM output probability serving as a scan-level deep learning score (DLS) as the pipeline output.

To validate the performance of the deep transfer learning technique, an additional independent test set composed of 4 CTProg and 12 NCTProg scans was evaluated by using the SVM trained with all 54 scans.
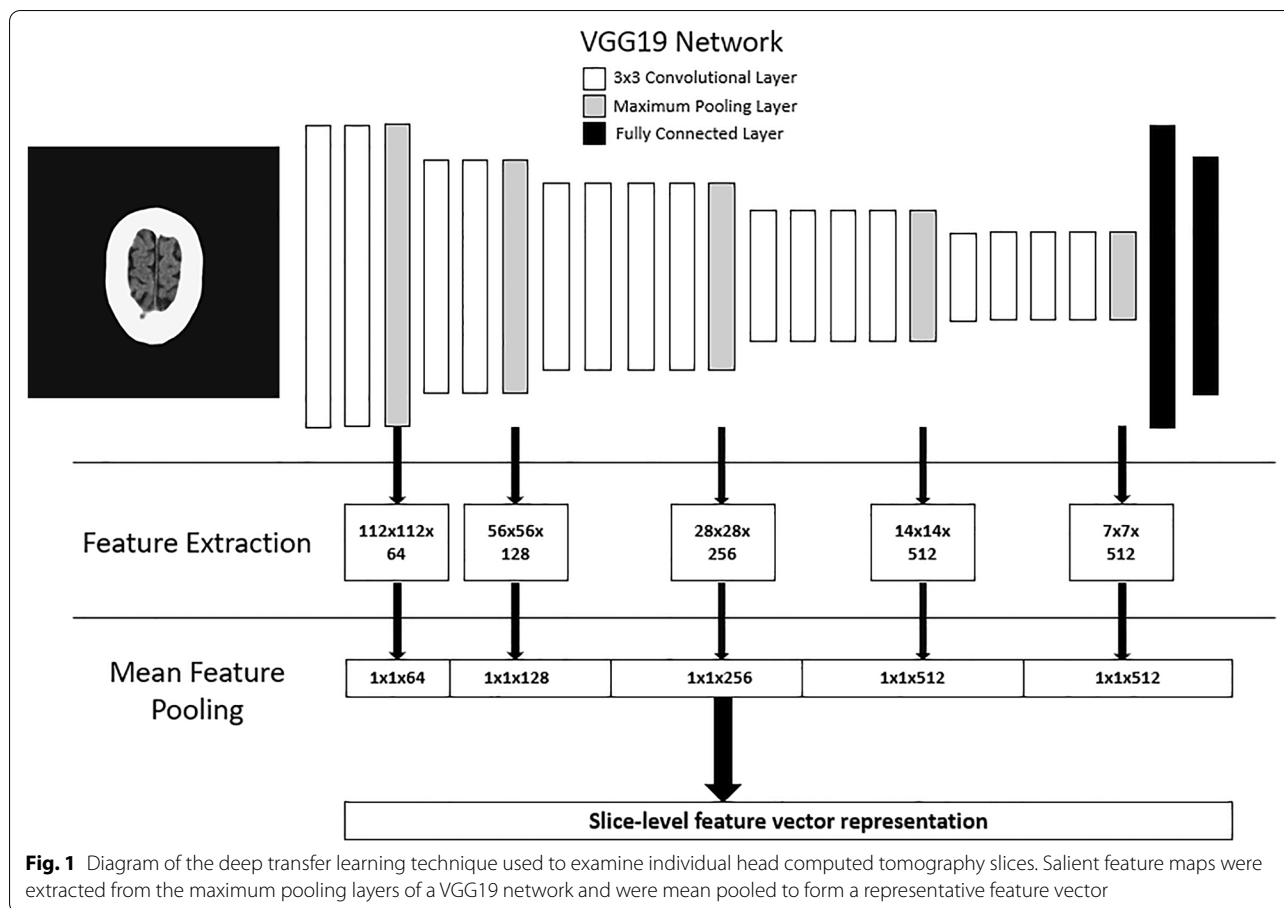
Statistical Analysis

Descriptive statistics were presented as means with standard deviations or medians with interquartile ranges (IQRs) (as appropriate) for continuous variables and as percentages for categorical variables. In univariate analyses, categorical variables were compared by using Fisher's exact test. A significance level was set at $p < 0.05$. All analyses were performed with the use of Python programming language (Python Software Foundation, https://www.python.org/) and R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

SVM classification performance between patients with CTProg and NCTProg was evaluated with receiver operating characteristic (ROC) curve analysis, with the area under the ROC curve (AUC) as the figure of merit. AUC confidence intervals (CIs) were determined through 1,000 bootstrapping resampling iterations. Note that because of the relatively small amount of data, some bootstrap iterations of the validation set only contained NCTProg and were thus ignored.

Results

Basic Characteristics of the Population

The basic characteristics of the population are summarized in Table 1. Overall, 54 patients were included in the

Fig. 1 Diagram of the deep transfer learning technique used to examine individual head computed tomography slices. Salient feature maps were extracted from the maximum pooling layers of a VGG19 network and were mean pooled to form a representative feature vector

analysis. The median age (IQR) of our cohort was 61 (16) years, and 25 patients (46%) were female. The predominant race was Black, with 44 patients (81%). The median time to achieving ROSC was 22 (23) minutes (Table 1).

Clinical Features on Initial Neurological Assessment
Among the 54 patients, the median GCS score (IQR) was 3 (3). At least one reactive pupil was appreciated in 43 patients (80%). At least one corneal reflex was appreciated in 29 patients (56%). A cough or gag reflex was present in 34 patients (64%), and an oculocephalic reflex was appreciated in 26 patients (52%). Thirty-seven patients (68%) were breathing over the ventilator. The median motor component of the GCS score was 1 (no motor response). Finally, 30 patients (56%) suffered from myoclonus (Table 1). The median time from arrest to initial HCT scan in the entire cohort was 3 (2–17) hours. The median time between the first and follow-up HCT scan was 2 (2–7) days.

Comparing CTProg and NCTProg Cohorts
Twenty-nine of the 54 patients had a follow-up HCT scan that demonstrated progression (CTProg), compared with 25 patients who did not (NCTProg) (Table 1). The median time from cardiac arrest to initial HCT scan and the time between first and follow-up HCT scans was not different across the two groups (Mann–Whitney U-test, $p = 0.408$ and $p = 0.398$, respectively). The CTProg group had a median GCS score of 3 (2), compared with 6 (4) in the NCTProg group ($p = 0.011$). Targeted temperature management was done in 27 (93%) patients with CTProg, compared with 17 (68%) patients with NCTProg ($p = 0.018$) (Table 1).

Mortality and Mechanism of Death
Of the 54 patients, 34 (63%) died in-house. Those included 12 (48%) patients whose CT imaging results did not progress and 22 (76%) whose CT imaging results progressed serially ($p = 0.035$).

The mechanism of death was withdrawal of life-sustaining therapies in 18 (82%) patients with CTProg, compared with 12 (100%) patients with NCTProg ($p = 0.107$). Cardiac death occurred in 2 (9%) patients with CTProg, compared with 0 (0%) patients with NCTProg ($p = 0.27$). Two (9%) patients with CTProg were declared brain
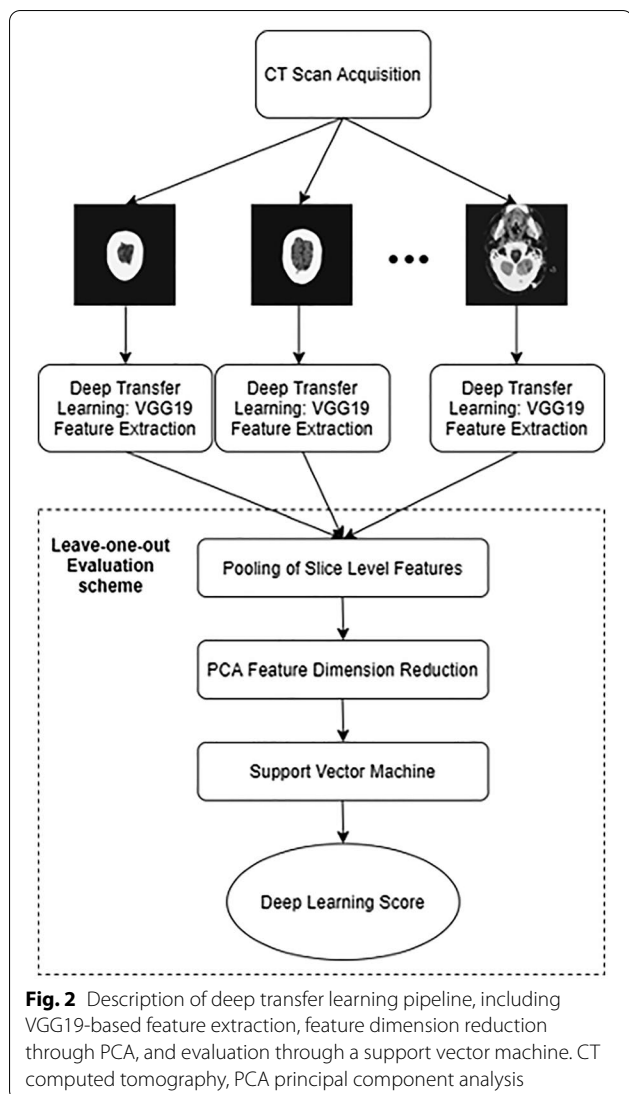
Fig. 2 Description of deep transfer learning pipeline, including VGG19-based feature extraction, feature dimension reduction through PCA, and evaluation through a support vector machine. CT computed tomography, PCA principal component analysis

dead, compared with 0 (0%) patients with NCTProg ($p = 0.27$) (Table 1).

Assessment of the Machine Learning Algorthm

In the task of distinguishing between patients with CTProg and NCTProg, the AUC was 0.96 (95% CI 0.91–1.0). The ROC curve is shown in Fig. 3 [14]. The prevalence-scaled optimal operating point of the ROC curve was found to be a DLS of 0.494. Operating at this threshold, performance included a sensitivity of 1.00, specificity of 0.88, accuracy of 0.94, and positive predictive value of 0.91. In evaluating the additional test set of 16 scans (4 CTProg and 12 NCTProg) and operating at the same DLS threshold of 0.494, the AUC was 0.90 (95% CI 0.74–1.00) (Fig. 3), with optimal operating performances

(sensitivity = 1.00, specificity = 0.66, accuracy = 0.75, and positive predictive value = 0.5).

Comparing Clinical Variables with DLS in HIBI Prediction

Table 1 shows that the difference in DLS is more significant than that of other variables. In a multiple variable logistic regression assessing DLS, pupillary reactivity, corneal reaction, VOR, and GCS score, DLS is the only variable significantly associated with progression (coefficient 275.35 [95% CI 106.826–443.875], $p < 0.01$).

Discussion

In this single-center study, we demonstrate that deep transfer learning can accurately identify a HCT signature of HIBI within the first 3 h after ROSC in comatose survivors of a cardiac arrest.

The determination and quantification of HIBI is a cornerstone of neuroprognostication in survivors of cardiac arrest. Similarly, it plays a determinant role in the shared decision-making process that often culminates in withdrawal of life-sustaining therapies in this patient population.

HCT is routinely used as part of this process. Absolute decrease in gray and white matter density, difference between gray and white matter density, and GWM ratio have been investigated in relation to neurological outcome. Although an absolute decrease in gray matter density alone is an unreliable predictor of poor neurological outcome, the predictive value increases when GWM ratio is considered. The main limitations to those studies include the time at which the analyzed HCT scan was performed (time from ROSC to HCT ranges anywhere from 4 to 72 h) and the choice of specific regions of interest within the brain to compare gray and white matter [9–12].

Our findings indicate that (1) a degree of identifiable injury to the brain may have already occurred in a number of patients who present normal-appearing findings on early HCT and (2) a significant number of patients presenting normal-appearing findings on HCT performed, on average, within the first 3 h after ROSC demonstrate significant abnormalities when HCT scans are evaluated with deep transfer learning.

Consideration for Early Distinction of Specific Endotypes of Cardiac Arrest Survivors

Our findings indicate that when early (within 3 h of ROSC) HCT scans with normal-appearing results are analyzed with deep transfer learning, two unique endotypes of cardiac arrest survivors can be identified: one type that bares no features of HIBI and one that does.

Table 1 Univariate analysis of CTProg and NCTProg groups

	Total (N = 54)	Progression (n = 29)	Nonprogression (n = 25)	p value
Median age (IQR), yrs	61 (16)	59 (27)	62 (13)	0.77
Female sex (%)	25 (46)	15 (52)	10 (40)	0.389
Race (%)				
African American	44 (81)	24 (83)	20 (80)	0.795
Whit	4 (7)	2 (7)	2 (8)	0.877
Asian	1 (2)	0 (0)	1 (4)	0.277
Unknown	5 (10)	3 (10)	2 (8)	0.767
Median GCS score (IQR)	3 (3)	3 (2)	6 (4)	0.011*
Median GCS-M (IQR)	1 (3)	1 (0)	3 (3)	0.034*
Pupillary reactivity (%)	43 (80)	20 (69)	23 (92)	0.036*
Presence of corneal reflex (%)	29 (56)	11 (41)	18 (72)	0.023*
Presence of VOR (%)	26 (52)	9 (35)	17 (71)	0.010*
Presence of gag/cough reflex (%)	34 (64)	15 (52)	19 (79)	0.038*
Spontaneous respiratory drive (%)	37 (68)	19 (65)	18 (72)	0.609
Myoclonus (%)	30 (56)	18 (62)	12 (48)	0.3
TTM (%)	44 (81)	27 (93)	17 (68)	0.018*
Median time to ROSC (min)	22 (23)	22 (20)	17 (15)	0.482
Median time to first CT (min)	163 (551)	138 (182)	220 (382)	0.408
Median time to second CT scan (min)	3,102 (3,011)	3,308 (2,863)	2,938 (3,825)	0.438
Median time between CT scans (min)	2,872 (2,991)	2,990 (2,303)	2,832 (2,787)	0.398
Mortality (%)	34 (63)	22 (76)	12 (48)	0.035*
Cause of death				
Cardiac death	2 (6)	2 (9)	0 (0)	0.27
Brain death	2 (6)	2 (9)	0 (0)	0.27
WLST	30 (88)	18 (82)	12 (100)	0.107
DLS	–	0.45 (0.02)	0.48 (0.01)	< 0.001*

DLS is the most significant discriminator between the two cohorts

CT computed tomography; *CTProg* CT progression; *DLS* deep learning score; *GCS* Glasgow Coma Scale; *GCS-M* Glasgow Coma Scale-Motor; *IQR* interquartile range; *NCTProg* no CT progression; *ROSC* return of spontaneous circulation; *TTM* targeted temperature management; *VOR* vestibulo-ocular reflex; *WLST* withdrawal of life-sustaining therapy

*$p < 0.05$

Because many of the therapies applied following ROSC intend to improve neurological outcomes, we suggest that the early stratification of survivors of cardiac arrest into these two distinct endotypes could serve to optimize the selection process of patients for clinical trials in the future. This model could help select patients who do not exhibit early radiographic HIBI because this endotype is, arguably, more likely to benefit from early interventions aimed at preventing further hypoxic-ischemic brain damage.

In addition, if our findings are prospectively reproduced and, at the same time, HCT progression (or presence of radiographic HIBI) is proven to definitively correlate with poor neurological outcome, we could also suggest that discussions about neurological prognosis with patients' surrogate decision makers could begin earlier than currently recommended.

Significance of Early Identification of Radiographic Brain Injury

The ability of the model to identify patients who will progress to radiographically discernable HIBI from those who will not, suggests that early on, changes on HCT already exist, albeit too subtle to discern by the human eye. Because substantial interobserver variability when identifying HIBI on HCT soon after out-of-hospital cardiac arrest has been reported [20], we verified that the CTProg and NCTProg groups were not evaluated by any one specific radiologist. As a matter of fact, a total of ten different radiologists composed the pool that evaluated the first scans. All those scans were read as normal, and the CTProg versus NCTProg determination was not more likely with any one particular radiologist.

The success of the automatic method in distinguishing patients with CTProg from those with NCTProg,
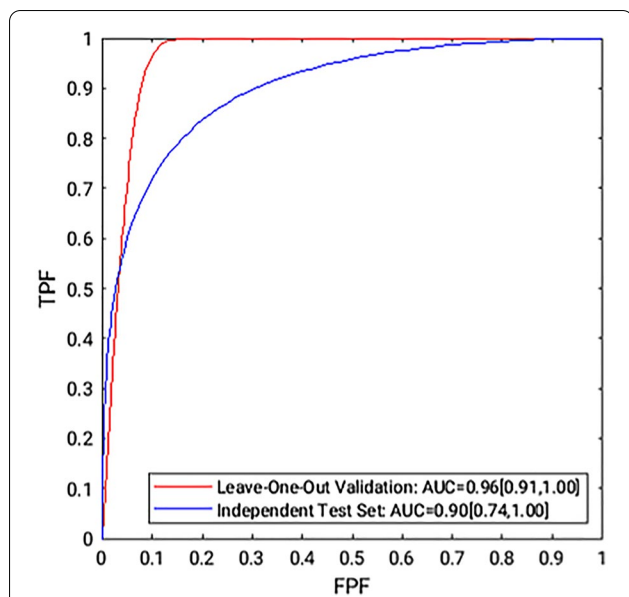
Fig. 3 ROC curves for the leave-one-out cross-validation approach and the independent test set in the task of distinguishing between patients with CTProg and NCTProg obtained by using a proper binormal model, with confidence intervals calculated through bootstrapping. AUC area under the ROC curve, CTProg computed tomography progression, FPF false positive fraction, NCTProg no computed tomography progression, ROC receiver operating characteristic, TPF true positive fraction

the degree of radiographic brain injury in follow-up HCT scans and therefore cannot comment on the potential discriminative power of the applied machine learning strategy in these two clinical scenarios. Furthermore, clinical characteristics, such as physical examination findings (GCS score, pupillary reactivity, and corneal reflexes), have been shown to correlate with neurological outcome, and although those variables are indeed different in our cohort of patients, the purpose of this study is to emphasize the unique role of early HCT imaging in identifying HIBI progression and not to replace the aforementioned clinical variables. To further understand the contribution of the DLS to prediction of HIBI progression, we conducted a pilot study using our clinical variables; a classifier based purely on the clinical variables without the DLS had an AUC of approximately 0.76, compared with a classifier based on the DLS alone, which had an AUC of ~0.96. Adding the clinical variables to the DLS in a combined classifier does not improve the AUC. That being said, a model incorporating clinical as well as radiographic features is out of the scope of the current work but is indeed the subject of future prospective research.

Our model does not predict mortality or withdrawal of life-sustaining therapies. It evaluates HCT images that are assessed as lacking signs of HIBI by the human eye and defines the cohort that will progress to show stigmata of HIBI on repeat imaging. In other words, it determines early on what HCT images bear features of HIBI that are not readily discernable by the human operator.

Limitations

This is a single-center study that will need prospective and multicenter validation. Also, the lack of a universally accepted radiographic definition of HIBI after cardiac arrest makes us rely on our neuroradiologists' assessments of HIBI on HCT in accordance with prior relevant literature [7, 8].

Additionally, because of the limited size of both the training and independent test data sets, there is a possibility for model overfitting and bias. Despite the promising results in both the leave-one-out-by-patient cross-validation technique and independent test set evaluation, a large diverse independent testing set is needed to further validate these results. The dynamic range of the DLS is small. There are multiple factors that could contribute to this phenomenon. First, the task itself indicates that the HCT images are similar, with all scans read as normal by a board-certified radiologist; thus, we expect that the embedded representations in feature space are clustered closely together, both within and between the NCTProg and CTProg populations. A second potential cause is the limited amount of data causing overfitting

particularly in the independent test set, suggests the presence of unique features within the images of the two cohorts. Given the presumed hypoxic-ischemic mechanism of brain injury, this finding can be interpreted in two different ways. The first is that HIBI might have already occurred to its full extent at the time of the initial HCT scan. If true, this would suggest that neuroprotective interventions started at this time could be of questionable utility. The injury would be potentially unmodifiable and determined by the clinical and temporal features of the cardiac arrest as well as the individual patient's profile.

On the other hand, it is possible that some patients have suffered some degree but not the full extent of hypoxic-ischemic injury. The injury is, therefore, a step along a potentially modifiable pathway. In other words, although a degree of injury might have occurred, its progression and outcome could be potentially modified by neuroprotective therapeutic measures. The model could help detect the patient population that is along this path and for which medical optimization may be more critical.

Although it is also possible that the early identification of brain injury with machine learning allows for a discrimination between different degrees of brain injury in the first few hours after ROSC, we have not quantified

in the model and thus introducing a biased evaluation. Although we have attempted to alleviate the concern of significant overfitting by demonstrating strong performance on the limited validation set and through CI estimation through bootstrapping, we acknowledge that there is still potential for a biased model. We are currently working to acquire a larger prospective data set for validation, but this lies outside the scope of this study. Given superior MRI sensitivity, it is indeed possible that some of the patients on presentation or follow-up HCT labeled as lacking radiographic signs of HIBI may have had HIBI on MRI. However, the particular focus of the study is to optimize CT interpretive potential and not claim any comparison or superiority to MRI.

Conclusions
Deep transfer learning reliably identifies HIBI in normal-appearing findings on HCT performed within 3 h after ROSC in comatose survivors of a cardiac arrest. This may suggest the presence of two distinct and identifiable endotypes of brain injury in this population, opening the door for more individualized treatment algorithms as well as providing a potential for early determination of neurological outcome. In addition to prospective validation, next steps of this work will include prospective patient cohorts with MRI and HCT imaging obtained and analyzed in tandem as well as incorporation of clinical variables into a combined clinical–imaging model.

Author details
[1] Neurosciences Intensive Care Unit, Department of Neurology, University of Chicago Medicine and Biological Sciences, 5841 S. Maryland Ave., MC 2030, Chicago, IL 60637-1470, USA. [2] Department of Neurological Surgery, University of Chicago Medicine and Biological Sciences, Chicago, IL, USA. [3] Department of Radiology, University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637-1470, USA.

Author contributions
AM, JDF, FDG, and MLG involved in study concept and design. FEA, JD, and AL involved in acquisition of data. AM, JDF, FDG, and MLG involved in analysis and interpretation of data. AM, JDF, FEA, FDG, and MLG involved in drafting the manuscript. AM, JDF, FDG, CL, CK, and MLG involved in critical revision of the manuscript for important intellectual content. AM, FEA, and JDF involved in statistical analysis.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Dragancea I, et al. The influence of induced hypothermia and delayed prognostication on the mode of death after cardiac arrest. Resuscitation. 2013;84(3):337–42.
2. Mulder M, et al. Awakening and withdrawal of life-sustaining treatment in cardiac arrest survivors treated with therapeutic hypothermia*. Crit Care Med. 2014;42(12):2493–9.
3. Dragancea I, et al. Protocol-driven neurological prognostication and withdrawal of life-sustaining therapy after cardiac arrest and targeted temperature management. Resuscitation. 2017;117:50–7.
4. Coute RA, et al. Disability-adjusted life years following adult out-of-hospital cardiac arrest in the United States. Circ Cardiovasc Qual Outcomes. 2019;12(3):e004677.
5. Geocadin RG, et al. Standards for studies of neurological prognostication in comatose survivors of cardiac arrest: a scientific statement from the American Heart Association. Circulation. 2019;140(9):517–42.
6. Rossetti AO, Rabinstein AA, Oddo M. Neurological prognostication of outcome in patients in coma after cardiac arrest. Lancet Neurol. 2016;15(6):597–609.
7. Sandroni C, D'Arrigo S, Nolan JP. Prognostication after cardiac arrest. Crit Care. 2018;22(1):150.
8. Moseby-Knappe M, et al. Head computed tomography for prognostication of poor outcome in comatose patients after cardiac arrest and targeted temperature management. Resuscitation. 2017;119:89–94.
9. Keijzer HM, et al. Brain imaging in comatose survivors of cardiac arrest: Pathophysiological correlates and prognostic properties. Resuscitation. 2018;133:124–36.
10. Yamamura H, et al. Head Computed Tomographic measurement as an early predictor of outcome in hypoxic-ischemic brain damage patients treated with hypothermia therapy. Scand J Trauma Resusc Emerg Med. 2013;21:37.
11. Choi SP, et al. The density ratio of grey to white matter on computed tomography as an early predictor of vegetative state or death after cardiac arrest. Emerg Med J. 2008;25(10):666–9.
12. Wang GN, et al. The prognostic value of gray-white matter ratio on brain computed tomography in adult comatose cardiac arrest survivors. J Chin Med Assoc. 2018;81(7):599–604.
13. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys. 2017;44(10):5162–71.
14. Giger ML. Machine learning in medical imaging. J Am Coll Radiol. 2018;15(3):512–20.
15. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. Acad Radiol. 2007;14(7):814–29.

16. Shin HC, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–98.

17. Karen Simonyan AZ (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR.

18. Horsch K, et al. A scaling transformation for classifier output based on likelihood ratio: applications to a CAD workstation for diagnosis of breast cancer. Med Phys. 2012;39(5):2787–804.

19. Horsch K, et al. Prevalence Scaling: Applications to an Intelligent Workstation for the Diagnosis of Breast Cancer. Acad Radiol. 2008;15:1446–57.

20. Caraganis A et al (2020) Interobserver variability in the recognition of hypoxic-ischemic brain injury on computed tomography soon after out-of-hospital cardiac arrest. Neurocrit Care